Molecular Modelling for Enzyme-Inhibition: A Search for a New Treatment for Cataract and New Antimicrobials and Herbicides

---

A thesis

submitted in partial fulfilment

of the requirements for the Degree of

Doctor of Philosophy in Biochemistry

at the University of Canterbury

Christchurch

New Zealand

Blair Gibb Stuart

March 2010

CONTENTS

ACKNOWLEDGEMENTS 1

ABSTRACT AND PUBLISHED WORK 2

1 INTRODUCTION 6

1.1 Calpain and the cataract hypothesis 6

1.2 Proteases 7

1.3 Calpains 10

1.4 Structure of the eye, cataract and the importance of an anti-cataract drug 14

1.5 The β-strand: important for protease recognition 15

1.6 Computer modelling programs 17

1.7 References 20

2 DEVELOPMENT OF A CALPAIN MODEL FOR DOCKING STUDIES 27

2.1 Introduction 27

2.1.1 Overview of calpain model development 27

2.2 Calpain X-ray crystal structures 28

2.2.1 The first published structures 28

2.2.2 Calpain constructs 1KXR and 1MDW 30

2.3 Exploring the calpain construct 1KXR to develop a viable model for Glide docking experiments 33

2.4 The InducedFit docking model 37

Contents

2.5 Conclusion 37

2.6 References 39

3 MOLECULAR MODELING OF ACYCLIC INHIBITORS 43

3.1 Introduction 43

3.1.1 Natural inhibitors 43

3.1.2 Modified natural inhibitors 45

3.1.3 Lead compound: SJA-6017 46

3.2 Docking studies of known inhibitors 46

3.2.1 Compounds of Inoue *et al* 46

3.2.2 Docking results for the Inoue *et al* compounds 50

3.3 Docking Studies of SJA-6017 analogues 58

3.3.1 *N*-Heterocyclic dipeptides 58

3.3.2 Docking results of *N*-heterocyclic dipeptides 60

3.4 Docking studies of diazo and triazene compounds 71

3.5 Conclusion 77

3.6 References 81

4 MOLECULAR MODELLING OF CYCLIC INHIBITORS 84

4.1 Introduction 84

4.2 First generation cyclic analogues: Modelling studies of 8-membered cyclic analogues of SJA-6017 84

4.3 Second generation cyclic analogues: modelling studies of ca 17 membered macrocyclic ring analogues of SJA-6017 91

4.3.1 Docking of macrocyclic analogues of SJA6017 using Glide 94

4.3.2 Docking of macrocyclic analogues of SJA6017 using InducedFit 101

4.3.3 Determination of Boltzmann weighted percentage of *β*-strand conformers within the macrocycle library of low energy ensembles 109

4.3.4 InducedFit docking of seven compounds with a >90% preference for a *β*-strand conformation 115

4.3.5 InducedFit docking: analogues of macrocyclic compounds A1, B1, C1, and H1 119

4.3.6 *In Vivo* study of compound 4.13 (CAT811) 129

4.3.7 Additional modification of macrocyclic cores: docking studies of macrocyclic diols 130

4.4 Conclusion 136

4.5 References 142

5 DRUG PENETRATION, *IN VIVO* TESTING, AND CRYSTALLOGRAPHY 146

5.1 Construction of a modified Ussing chamber 146

5.2 Testing of inhibitors using the Ussing chamber to determine corneal Penetration 148

5.3 *In-vivo* sheep trial results of compound **3.26** 150

5.4 Crystallography 154

5.5 Conclusion 155

5.6 References 157

6 DESIGN OF ANTIMICROBIALS AND HERBICIDES 158

6.1 Introduction 158

6.2 Modeling studies of dehydroquinase 159

6.3 Conclusion 164

6.4 References 166

7 MOLECULAR MODELLING TO DETERMINE THE ABSOLUTE CONFIGURATION OF CHRYSOSPORIDE 167

7.1 Introduction 167

7.2 Modelling 168

7.3 Conclusion 171

7.4 References 171

8 APPENDIX 172

8.1 Protocol 1: Refinement of the Glide model from 1KXR 172

8.2 Protocol 2: Refinement of the InducedFit model from 1KXR 172

8.3 Protocol 3: Conformational search methods 173

8.4 Protocol 4: Glide docking protocol 173

8.5 Protocol 5: InducedFit docking protocol 174

8.6 Protocol 6: Computational methods used for diazo- and triazene-dipeptide aldehydes in section 3.4 174

8.7 Protocol 7: Preparation of dehydroquinase model 175

8.8 Protocol 8: Glide docking protocol for dehydroquinase 175

8.9 Protocol 9: Bodipy assay 176

8.10 References 176

Acknowledgements

Firstly I would like to give special thanks to my supervisor Prof. James Coxon for all his advice and support through this long journey.

I would also like to thank my co-supervisor Prof. Andrew abell for his input into this thesis and my associate supervisor Dr. Quentin MacDonald for his help and advice with molecular modelling problems.

A big thanks goes to the Canterbury cataract research team for all their help. Thanks especially to Dr. Axel Neffe for his modelling expertise, Dr. Mathew Jones, Dr. Steve Aitken, and Dr. Steven McNabb for all their advice and synthesis of compounds, and Dr. Janna Nikkel who performed almost all of the *in vitro* testing.

Thanks also goes to the Lincoln cataract group headed by Dr. James Morton and Professor Roy Bickerstaffe. Collaboration with the Lincoln group allowed access to the equipment for the BODIPY *in vitro* test and members of this group performed the *in vivo* studies on sheep.

Most importantly I would like to thank all my friends and family for all their love and support through this process.

Abstract

Background information is given in chapter 1 on protease enzymes in general along with information specific to the calpain enzyme and its role in human disease including cataract. It also discusses the importance of the β-strand conformation for substrates that are recognized by protease enzymes. Almost all known protease substrates bind their respective targets in a β-strand. An overview of the modelling methodology and the modelling programs used in this thesis are discussed.

Descriptions of the most important X-ray crystal structures that have been downloaded from the Protein Database (PDB) and used in this thesis are discussed in chapter 2. This includes the X-ray structure 1KXR, a construct of μ-calpain that was the first structure to show the catalytic triad of a calpain in an active conformation. The elucidation of this structure was important as we have used it to develop a working model for the docking experiments central to the work presented here.

The results from docking experiments with numerous potential inhibitors are discussed in chapter 3, especially those from the work of Inoue *et al*. These include our lead compound SJA-6017 and analogues. The docking experiments on these compounds were performed to test our enzyme model and get 'baseline' data for subsequent docking of compounds to establish those members of our virtual library that may be worth synthesising. We also included data from docking experiments using acyclic compounds that are based on the lead compound and those synthesized by members of our cataract group. These include *N*-heterocyclic compounds, diazo- and triazene-dipeptides and are presented along with their IC_{50} and *in vivo* testing results.

Abstract

In chapter 4 data is presented from docking experiments performed on the macrocyclic compounds that were considered and those that have subsequently been synthesized in house by members of the cataract research team. The importance of the β-strand conformation is discussed and how the presence of the macrocycle can increase the propensity of the molecule in a bioactive conformation, in this case a *β*-strand. Many of the macrocycle compounds presented in this chapter are shown to favour a *β*-strand conformation and can dock into our enzyme model in such a way as to indicate they may be good inhibitors. The docking study is backed up by *in vitro* IC_{50} and *in vivo* test results.

To test the effectiveness of two of our best inhibitors at crossing the cornea (the major barrier to ocular delivery of an inhibitor to the eye lens) an apparatus called an Ussing chamber which I designed to fit our requirements.The chamber was expertly constructed by departmental workshop staff. The Ussing chamber was designed to fit a sheep cornea between two chambers so that corneal diffusion of an inhibitor could take place. The amount of inhibitor diffusing across the cornea at different time intervals was to be determined by HPLC or if facilities existed by radioactive labelling.One of our best *in vitro* inhibitors was also the subject of a sheep trial at Lincoln University (Morton and Bickerstaffe) and the results are given. An attempt to co-crystallize several of our inhibitors into papain (a very similar cysteine protease to calpain) was undertaken but without success.

In the two final chapters we report on two studies not involved with calpain:

In chapter 6 -dehydroquinase, an enzyme which catalyses the third step in the shikimate pathway and a sensible drug target for designing antibiotics and herbicides was examined by

molecular modelling. Docking studies were performed using PDB X-ray structures of dehydroquinase as the target enzyme. Inhibitors designed and synthesized in house by Dr Mary Gower in her PhD were docked into a model of dehydroquinase to study how they potentially bind the active site of the enzyme. The docking data is compared with *in vitro* testing results.

In the final chapter we report a study of a cyclic pentapeptide, chrysosporide, isolated from a New Zealand fungus by bioactivity-guided fractionation. The planar structure was deduced by detailed spectroscopic analysis, and the absolute configurations of the amino acid residues were defined by Marfey's method. As both enantiomers of Leu occurred in chrysosporide, molecular mechanics calculations were applied to the analysis to distinguish between the possible structural isomers. Only the lowest energy conformers of the cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu) isomer could explain the observed NOEs, suggesting that this was the most probable amino acid sequence for chrysosporide.

Published work from this thesis:

Papers

1. Abell, A. D.; Jones, M. A.; Coxon, J. M.; Morton, J. D.; Steven G. Aitken; McNabb, S. B.; Lee, H. Y.-Y.; Mehrtens, J. M.; Alexander, N. A.; Stuart, B. G.; Neffe, A. T.; Bickerstaffe, R., Molecular Modeling, Synthesis, and Biological Evaluation of Macrocyclic Calpain Inhibitors. *Angew. Chem. Int. Ed.* **2009,** *48*, 1455 -1458.

2. Abell, A. D.; Jones, M. A.; Neffe, A. T.; Aitken, S. G.; Cain, T. P.; Payne, R. J.; McNabb, S. B.; Coxon, J. M.; Stuart, B. G.; Pearson, D.; Lee, H. Y.-Y.; Morton, J. D., Investigation into the P3 Binding Domain of m-Calpain Using Photoswitchable Diazo- and Triazene-dipeptide Aldehydes: New Anticataract Agents. *J. Med. Chem.* **2009,** *50*, 2916-2920.

3. Abell, A. D.; Jones, M. A.; Neffe, A. T.; Steven G. Aitken; Cain, T. P.; Payne, R. J.; McNabb, S. B.; James M. Coxon; Stuart, B. G.; Pearson, D.; Lee, H. Y.-Y.; Morton, J. D., Investigation into the P3 Binding Domain of m-Calpain Using Photoswitchable

Diazo- and Triazene-dipeptide Aldehydes: New Anticataract Agents. *J. Med. Chem.* **2007,** *50*, 2916-2920.

4. Abell, A. D.; Brown, K. M.; Coxon, J. M.; Jones, M. A.; Miyamoto, S.; Neffe, A. T.; Nikkel, J. M.; Stuart, B. G., Synthesis and evaluation of eight-membered cyclic pseudo-dipeptides. *Peptides* **2005,** *26*, 251-258.

5. Mitova, M. I.; Stuart, B. G.; Cao, G. H.; Blunt, J. W.; Cole, A. L. J.; Munro, M. H. G., Chrysosporide, a Cyclic Pentapeptide from a New Zealand Sample of the Fungus *Sepedonium chrysospermum*. *J. Nat. Prod.* **2006,** *69*, 1481-1484.

Patents

1. Abell, A. D.; Coxon, J. M.; Jones, M. A.; McNabb, S. B.; Neffe, A. T.; Aitken, S. G.; Stuart, B. G.; Nikkel, J. M.; Duncan, J. K.; Klanchantra, M.; Morton, J. D.; Bickerstaffe, R.; Robertson, L. J. G.; Lee, H. Y. Y.; Muir, M. S. Preparation of macrocyclic peptide alcohols and aldehydes as calpain inhibitors and their compositions, PCT Int. Appl., WO 2008048121, **2008**.

2. Abell, A. D.; Coxon, J. M.; Miyamoto, S.; Jones, M. A.; Neffe, A. T.; Aitken, S. G.; Stuart, B. G.; Nikkel, J. M.; Morton, J. D.; Bickerstaffe, R.; Robertson, L. J. G.; Lee, H. Y.; Muir, M. S. Preparation of peptide alcohols and aldehydes as calpain inhibitors,N.Z. Patent. **2007**, 547303; U.S. Patent Appl., **2007**, 20070293560.

1 Introduction

1.1 Calpain and the cataract hypothesis

There have been several reports[1-10] that cataract development results from unregulated Ca^{2+} mediated degradation of lens crystallins. The calpain isoform m-calpain, a cysteine protease, is known to be a major player in cataract formation in rodent lenses and recent evidence indicates that over-activation by Ca^{2+} causes cataractogenesis in other mammals.[2]

m-Calpain has been isolated from human lenses[2] and it is reasonable to hypothesize that elevated levels of Ca^{2+} can result from damage to the eye, such as ultraviolet light exposure, thereby, leading to over-activation of m-calpain and an increase in proteolysis of lens crystallins at specific protein sites.[2]

The proteins in the eye lens, namely crystallins, are extremely soluble proteins even at the high concentrations found in the eye lens and form a clear crystalline array resulting from specific protein-protein packing.[3-10] Any degradation of the crystallins will disrupt their crystalline packing and subsequent precipitation can cause formation of cataract.[3-10]

Since over-activation of m-calpain is involved in the hydrolysis of crystallins and thereby cataract formation it is, therefore, a drug target. Inhibition of this enzyme is expected to retard cataract formation.[2]

Calpains have also been implicated in many other degenerative diseases including Alzheimer's and muscular dystrophy.[11-15]

1.2 Proteases

Proteases (peptidases) are a large group of enzymes that hydrolyze peptide bonds of proteins and peptides. These are classified as being either endopeptidases, which hydrolyze the internal bonds of proteins or peptides, or exopeptidases, which hydrolyze terminal peptide bonds.[16]

The human genome contains 23 pairs of chromosomes containing about 3,000,000,000 base pairs. The genome contains about 35,000 genes, each gene coding for a gene product. Proteases account for about 700 (2%) of these gene products and some 100 (14%) of these are under active investigation as drug targets.[17]

There are currently six types of proteases recognised; aspartic, cysteine, glutamic, serine, threonine, and metallo-. The first five of these are characterised by the amino acid in the active site of the enzyme, and the sixth, the metallo-proteases, by the presence of a bivalent metal ion at the active site. The metal is usually a zinc ion, but can be cobalt, iron or manganese.[16]

The scissile bond (peptide cleavage site) defines the starting point for protease active site nomenclature. The enzyme subsites are numbered S_1, S_2, S_3 (S for subsite) and so on from the scissile bond towards the N-terminus, and S_1', S_2', S_3'… moving outward to the C-terminus. Similarly, the amino acid residues of the peptide substrate are numbered P_1, P_2, P_3, (P for peptide) and P_1', P_2', P_3'.[16] This nomenclature is outlined in **Figure 1.1**. The S subsites are critical points in determining the peptide substrate specificity of each protease. Subsites have

preferences for different peptide (P) substrate amino acids. Changes to P substrate amino acids usually alter the enzyme-substrate binding affinity.

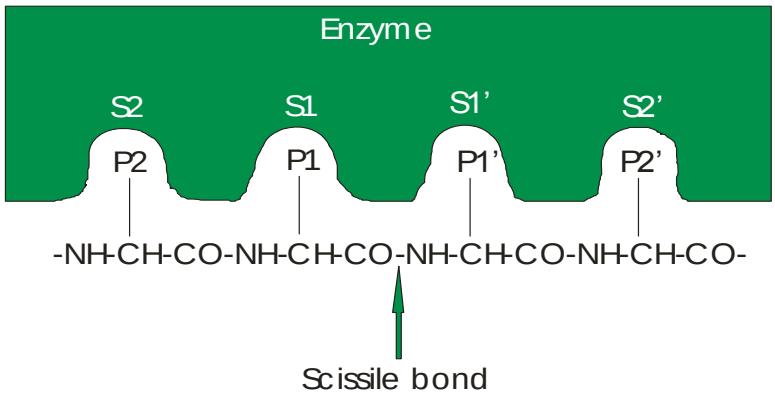


Figure 1.1: A typical protease cleavage site with bound peptide substrate showing the non-prime side (S) at the N-terminus and the prime side (S') at the C-terminus. Adapted from Hooper 2002.[16]

Research has shown proteases to be good targets for therapeutic drug discovery because they are involved in key metabolic processes. These include regulating the cell cycle, cell growth, cytoskeletal remodelling, antigen processing and apoptosis (programmed cell death).[16, 18] When these processes do not function properly it can result in diseases such as Alzheimer's and cataract.[18, 19] In addition to protease targets involved in human metabolism, there are known protease targets of disease-causing parasites, bacteria, and viruses such as malaria, anthrax, and HIV, respectively.

Examples of aspartic protease targets are renin and HIV which are important enzyme drug targets for the control of hypertension and HIV, respectively.[20, 21]

The cysteine proteases have been problematic for drug design, as many of their known inhibitors contain an electrophilic "warhead" that binds covalently in a reversible or irreversible manner making the inhibitor particularily reactive. This reactivity means that they typically bind non-specifically to other proteins, receptors and other biomolecules *in vivo*.

This can give rise to problems with toxicity, drug metabolism and other pharmacokinetics (the branch of pharmacology that studies the fate of pharmacological substances in the body, as their absorption, distribution, metabolism, and elimination).[22, 23] However, there is considerable interest in developing selective inhibitors that address these problems. Calpains, which are a class of cysteine protease and the focus of this thesis, are implicated in many degenerative diseases including Alzheimer's, muscular dystrophy, and cataract.[11, 12, 14, 15, 18, 19, 23]

Glutamic proteases, once thought to be part of the aspartic proteases, are a recently re-classified catalytic type that has, so-far, only been isolated from fungal species of Ascomycota.[24]

In recent years metallo-protease drug targets have included those found in the Malaria parasite *Plasmodium falciparum*,[25, 26] those targeted to treat cancer and arthritis,[27] and even a metallo-protease from the anthrax causing bacteria *Bacillus anthracis*.[28]

The most studied of the proteolytic enzymes are the serine proteases. Typical serine protease targets are thrombin to treat thrombosis,[29] viral enzyme NS3-protease to treat Hepatitis C,[30] and DPP IV to treat diabetes.[31]

The threonine proteases include proteasome. Inhibitors of proteasome are potential anti-cancer drugs interacting with the proteasome-ubiquitin pathway.[32]

1.3 Calpains

Calpains are a group of proteases belonging to the cysteine protease group of proteolytic enzymes. Strictly speaking, they belong to one of at least seven different evolutionary distinct groups of cysteine proteases.[33] Some examples of these distinct groups, known as clans, are shown in **Table 1.1**.

Table 1.1: Clans and Families of Cysteine proteases (adapted from Barrett and Rawlings, 2001).[33]

Clan	Family	Type example	PDB Structure
CA	C1	papain (*Carica papaya*)	1PE6
	C2	calpain (*Gallus gallus*)	1DFO
	C27	Leader proteinase (foot-and-mouth disease virus)	1QMY
CD	C14	Caspase-1(*Rattus norvegicus*)	1ICE
	C25	Gingipain R (*porphyromonas gingivalis*)	1CVR
CE	C5	Adenain (human adenovirus type 2)	1AVP
	C48	Ulp1 endopeptidase (*Saccharomyces cerevisiae*)	1EUV
CF	C15	Pyroglutamyl-peptidase 1 (*Bacillus amyloliquefaciens*)	1AUG
CH	C46	Hedgehog protein (*Drosophilia melanogaster*)	1AT0
PA	C3	Picornain 3C (poliovirus type 1)	1HAV
	C37	Processing peptidase (Southhampton virus)	
PB	C44	Glutamine phosphoribosylpyrophosphate amidotransferase precursor (*Homo sapiens*)	1GPH

1 Introduction

Calpains are intracellular proteolytic enzymes that are activated by Ca^{2+} [11-13] and found in the cytosol of almost all mammalian cells and in other animals and fungi. Calpains belong to the largest clan of cysteine proteases, clan CA, a clan that contains half of all cysteine protease families. Cysteine proteases typically have an active site consisting of a catalytic dyad containing a cysteine and a histidine.[33] Calpain itself has a catalytic triad that contains both the cysteine and histidine with the addition of an asparagine.

In mammals there are various isoforms found, some are tissue specific, others have been isolated from almost all tissue types. Some tissue specific calpains include; p94 – skeletal muscle, Lp82 – lens, nCL-2 – stomach, nCL-4 – digestive organ, CAPN6 – placenta.

The ubiquitous calpains known as μ- and m-calpain are also known as calpain I and calpain II, respectively, and both are well characterised.[2] They require micromolar and millimolar Ca^{2+} concentrations, *in vitro*, for activation, respectively.

m-Calpain is a heterodimer consisting of six domains and this isoform is one of the best characterized calpains. The closely related and similarly well characterised isoform, μ-calpain, [11, 12, 15] and a construct made up of domains IIa and IIb of each of these calpains (μ-calpain and m-calpain) have provided a valuable model for our studies.[13, 18]

Both μ- and m-calpain are composed of two subunits (**Figure 1.2**), a large 80kD catalytic subunit (orange), and a small 30kD regulatory subunit (blue). The large subunit is made up of domains I to IV and the smaller subunit of domains V and VI. Calcium binding at DIV and DVI is required for activation as this brings the catalytic triad of calpain in the active site (red star) located at DII into the active proteolytic conformation.[2]

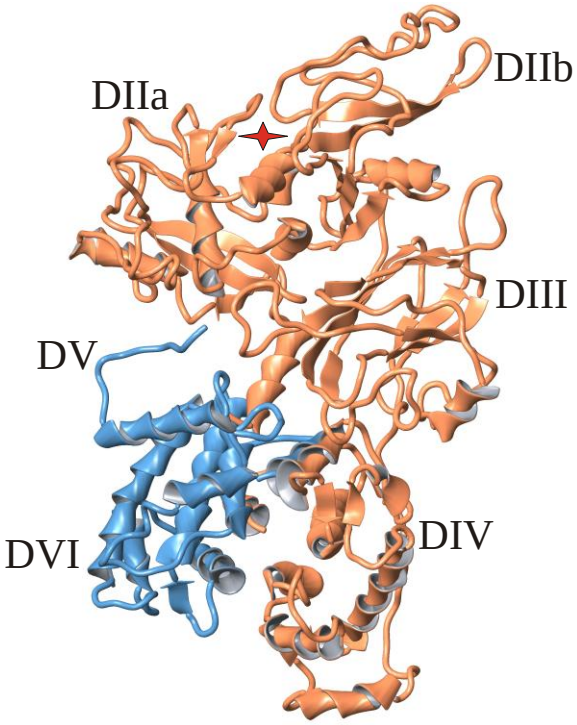


Figure 1.2: Schematic domain representation of m-calpain. Domains I – IV make up the large 80kD (orange) subunit and domains V and VI comprise the small 30kD subunit (blue). Domain II is split into IIa and IIb which represents the two sides of the active site cleft (red star). Calcium binding occurs in domains IV and VI which facilitates the activation of the enzyme.

Calcium activation causes a large and significant change in the conformation of the active site, a change that results in the arrangement of a catalytic triad in a conformation essential for enzyme activity.[2] This is depicted in **Figure 1.3**.

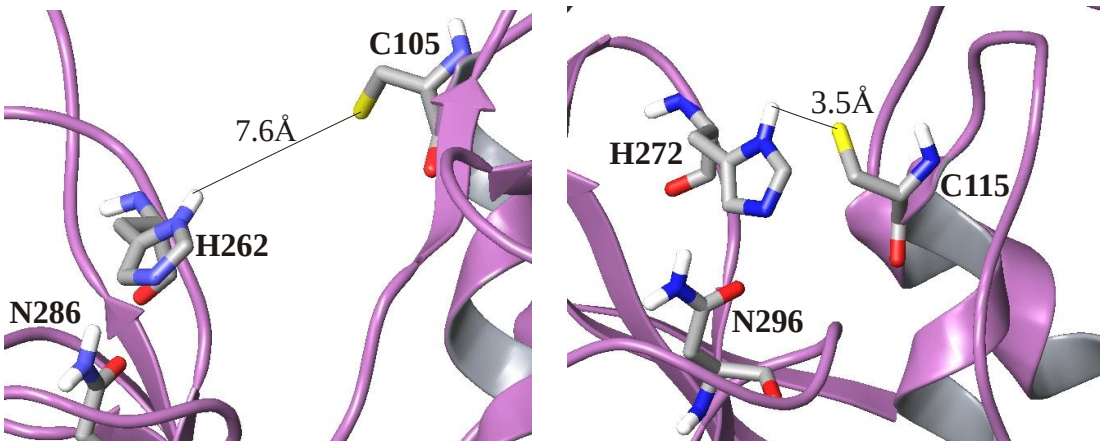


Figure 1.3: Left, 1KFU, a crystal structure of calpain not in the active conformation showing a catalytic triad with a distance of 7.6Å between the cysteine and histidine. Right, 1KXR, a crystal structure of μ-

calpain that is in the active conformation showing a catalytic triad with a distance of 3.5Å between the cysteine and histidine.

The mechanism by which the enzyme cleaves peptide substrates is shown in **Figure 1.4** below. As described above, the enzyme consists of a catalytic triad which facilitates the hydrolysis of amide bonds. Firstly the enzyme forms a thioacetyl intermediate which is subsequently cleaved by an activated water molecule to leave the cysteine in its free thiol state ready to cleave another amide.

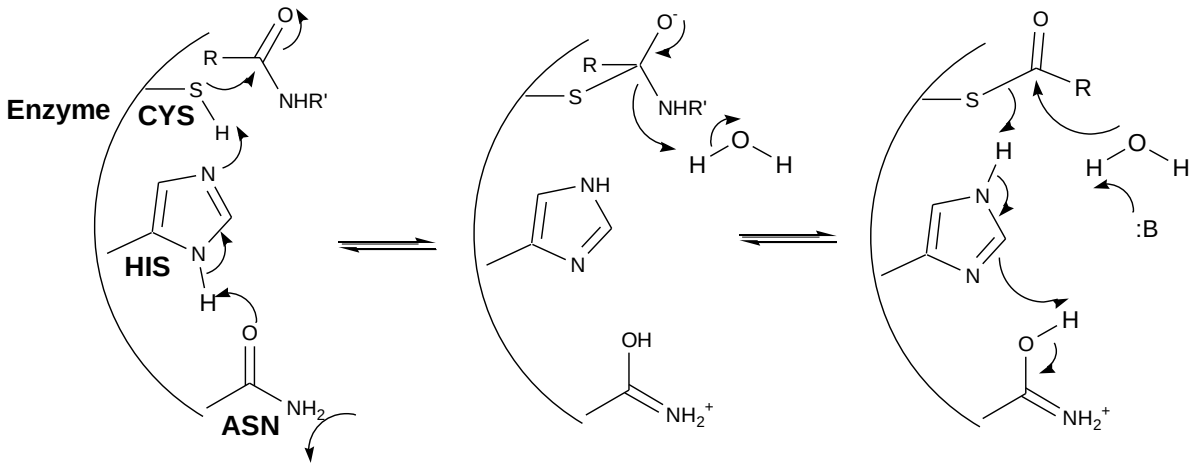


Figure 1.4: Schematic of Cysteine protease mechanism.

Calpains have specific subsite preferences for certain amino acid residues. The μ- and m-calpains show high structural homology and, therefore, have similar subsite preferences. Several reviews[34-36] on calpain inhibitors report a preference at the P_1 position for leucine, the P_2 position prefers a valine or leucine, and at the P_3 position the presence of large aromatic residues is favourable.

1.4 Structure of the eye, cataract and the importance of an anti-cataract drug

The structure of the human eye in **Figure 1.5** shows a cross-section with the location of the lens in relation to the other major parts of the eye. The lens is a specialized non-vascular tissue that focuses an image onto the retina which then triggers electrical impulses that are interpreted by the brain as vision.[2] In the lens are protein molecules called crystallins that form a transparent crystalline array. Because the cells of the lens have no protein synthetic or degradation machinery the lens crystallins need to last a lifetime.[4] As a consequence, any damage to the lens is usually permanent. Damage to the lens can occur by factors such as UV light and free radicals which, as described earlier, can cause an influx of Ca^{2+} into the lens. This over-activates calpain which, in turn, begins to hydrolyse the lens crystallins causing opacification. In cataract, opacification of all or part of the lens prevents or distorts visible light from reaching the retina, thereby reducing optical performance.[2]

Age related cataract in the developed world can be easily corrected by an operation to remove the affected lens and replace it with an artificial lens. However, waiting lists for this operation can be long with quality of life suffering. Therefore, the development of a topically applied eye treatment in the form of an eye drop to retard the progression of cataract would be an important alternative to surgery.

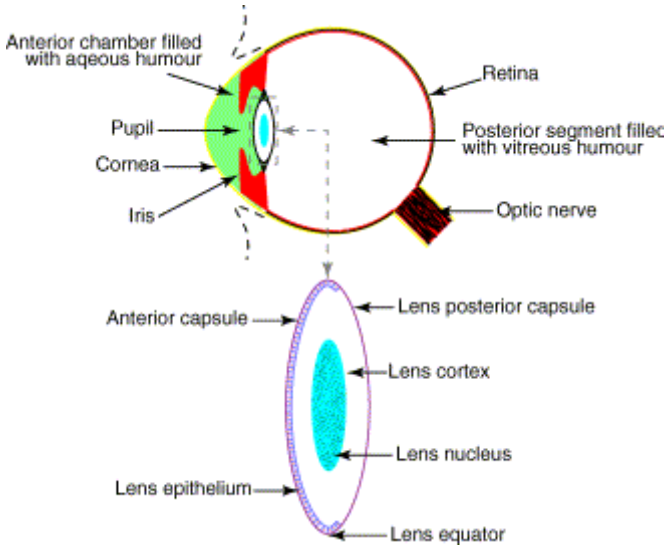


Figure 1.5. Schematic cross-section of the human eye showing the location of the lens.[2]

In the undeveloped world surgery to treat cataract is virtually non-existant and sufferers spend their latter years in a blind state. A cheap eye drop treatment to retard cataract in these countries would prevent a lot of suffering.

1.5 The β-strand: important for protease recognition

It has been reported in the literature and this study has confirmed that protease substrates, including most peptidomimetic inhibitors of proteases, bind in a *β*-type strand conformation in the region of the active site.[37-41]

The X-ray crystal structures (1TL9 and 1TL0) of truncated calpain enzymes, co-crystallized with two well known calpain inhibitors, have been deposited in the Protein Database (PDB).[42, 43] These and many other X-ray crystal structures of proteases found in the ever expanding PDB support the assumption that the inhibitors bind the active site in a *β*-

strand conformation. This conformation facilitates the formation of several hydrogen bonds, and results in the formation of a mini β-sheet with active site residues (**Figure 1.6** and **1.7**).

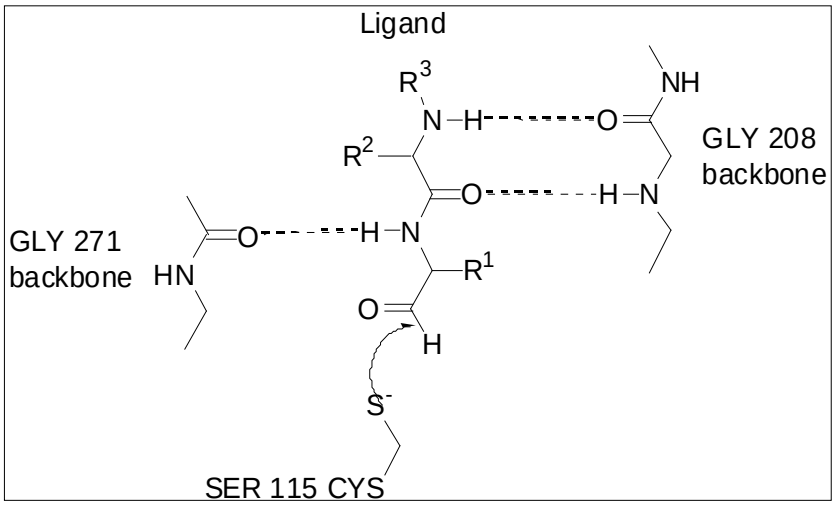
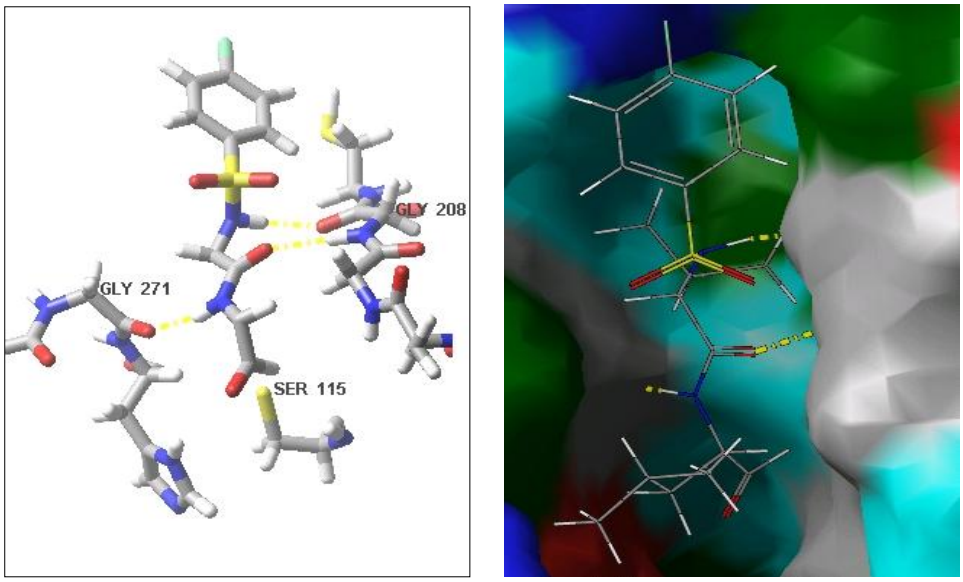


Figure 1.6: Diagram of a typical peptidomimetic ligand docked in the active site of mini μ-calpain according to the X-ray structure (PDB 1KXR) before attack from CYS 115. The formation of β-sheet type H-bonding between the ligand and the backbone of two GLY residues seems to be a requirement of good binding and, therefore, good inhibition. In addition, the electrophilic end (e.g. the aldehyde) needs to be in a position for nucleophilic attack by the S of CYS 115, typically within 5 Å.

These H-bonds are formed between the ligand and two glycine residues in the active site (Gly208 and Gly271) and are important for stabilisation of the bound ligand. Bound in this way the ligand's warhead is in a position for nucleophilic attack by the active site cysteine (Cys115).



A) B)

Figure 1.7: A) Picture of a typical peptidomimetic ligand bound to active site of X-ray structure of mini μ-calpain (PDB 1KXR) before attack from CYS 115. Note the amino acid side chains of the ligand have been omitted for clarity. B) A surface picture of the same X-ray structure with ligand bound.

1.6 Computer modelling programs

The computer-based rational design of new therapeutic agents is now possible and popular because of the availability of faster and cheaper computers, state of the art effective software, and a plethora of enzyme X-ray structures (potential drug targets).

Molecular mechanics programs such as Macromodel[44-48] are extremely useful and indeed such programs are crucial for rational drug design and development.[49] They can first be used to search the conformational space of potential drug candidates. The ensemble of low energy conformers generated by a conformational search can then be subjected to cluster analysis, performed by programs such as Xcluster.[50, 51] X-cluster has been written by Dr D. Quentin McDonald in conjunction with Schrodinger. Quentin is a former graduate of our Canterbury modeling group. His program greatly reduces the number of distinct conformers one has to

deal with by sorting closely related conformers into clusters. Representative conformers (determined by the Xcluster program) from each cluster can then be analyzed visually and statistically against conformations (computational or X-ray) of known inhibitors for their drug potential. MacroModel has proved to be reliable in reproducing the bioactive conformations of many enzyme inhibitors.[49]

A particularly useful advent is that of automated docking programs, which use existing X-ray structure coordinates of an enzyme to explore how a potential inhibitor might bind.[52-54] The information obtained from these experiments give clues about which inhibitors could bind to the enzyme best, and thus, which ones should be synthesized. Glide is an automated docking program that allows small ligands to be flexibly and optimally docked to a protein/enzymes 'active site'.[55] Each docked ligand has a score calculated based on specific parameters and a specified number of best fit conformers is kept for each ligand. The best fit conformers, called “poses”, can also be analyzed visually and statistically (superimposition using root mean square deviations of atomic coordinates) and compared to any enzyme-ligand co-crystallized structures that are available. This can be valuable for determining any features of binding that may be essential for an inhibitor to be effective at binding the enzyme.

A potential flaw in automated docking programs is that, while the ligands are treated as flexible molecules, the protein is treated as a rigid entity. It is, however, well known that many enzymes and receptors have somewhat flexible active sites in order to accommodate their respective substrates. It is well accepted that the Ca^{2+} activated calpains undergo considerable and important changes in their conformation upon Ca^{2+} binding and further changes occur in the active site when a substrate binds.[12, 13, 18, 56] When the software

accommodates this, it is known as an induced fit docking protocol and Schrodinger's InducedFit Docking Protocol is used extensively in this thesis.[57, 58]

The InducedFit Docking Protocol uses a combination of Glide[59] and a protein structure prediction, preparation and refinement program called Prime.[60] An InducedFit Docking job is actually a series of Glide and Prime jobs that simulate the flexible movement of both enzyme and ligand.

1.7 References

1. Fukiage, C.; Nakajima, E.; Ma, H.; Azuma, M.; Shearer, R., Characterization and Regulation of Lens-specific Calpain Lp82. *J. Biol. Chem.* **2002,** *277*, 20678-20685.
2. Biswas, S.; Harris, F.; Dennison, S.; Singh, J.; Phoenix, D. A., Calpains: Targets of Cataract Prevention? . *TRENDS in Molecular Medicine* **2004,** *10*, 78-84.
3. Basak, A.; Bateman, O.; Slingsby, C.; Pande, A.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J., High-resolution X-ray Crystal Structures of Human gammaD Associated with Aculeiform Cataract. *J. Mol. Biol.* **2003,** *328*, 1137-1147.
4. A. G. Purkiss, O. B., J. M. Goodfellow, N. H. Lubsen, C. Slingsby. (2002). The X-ray Crystal Structure of Human gammaS-crystallin C-terminal Domain. J. Biol. Chem. 277 4199-4205.; Bateman, O.; Goodfellow, J. M.; Lubsen, N. H.; Slingsby, C., The X-ray Crystal Structure of Human gammaS-crystallin C-terminal Domain. *J. Biol. Chem.* **2002,** *277*, 4199-4205.
5. Lampi, K. J.; Ma, Z.; Hanson, S. R. A.; Azuma, M.; Shih, M.; Shearer, T. R.; Smith, D. L.; Smith, J. B.; David, L. L., Age related Changes in Human Lens Crystallins Identified by Two-dimensional Electrophoresis and Mass Spectrometry. *Exp. Eye Res.* **1998,** *67*, 31-43.
6. Srivastava, O. P.; Srivastava, K., Degradation of gammaD- and gammaS-Crystallins in Human Lenses. *Biochem. Biophys. Res. Comm.* **1998,** *253*, 288-294.
7. Srivastava, O. P.; Srivastava, K., Beta B2-crystallin Undergoes Extensive Truncation During Aging in Human Lenses. *Biochem. Biophys. Res. Comm.* **2003,** *301*, 44-49.
8. Hanson, S. R. A.; Hasan, A.; Smith, D. L.; Smith, J. B., The Major In Vivo Modifications of the Human Water- insoluble Lens Crystallins are Disulfide Bonds,

Deamidation, Methionine Oxidation and Backbone Cleavage. *Exp. Eye Res.* **2000,** *71*, 195-207.

9. Ueda, Y.; Fukiage, C.; Shih, M.; Shearer, T. R.; David, L. L., Mass Measurements of C-terminally Truncated alpha-Crystallins from Two-dimensional Gels Identify Lp82 as a Major Endopeptidase in Rat Lens. *Molecular and Cellular Proteomics* **2002,** *1.5*, 357-365.

10. Lund, A. L.; Smith, J. B.; Smith, D. L., Modifications of the Water-insoluble Human Lens alpha-Crystallins. *Exp. Eye Res.* **1996,** *63*, 661-672.

11. Todd, B.; Moore, D.; Deivanayagam, C. C. S.; Lin, G.; Chattopadhyay, D.; Maki, M.; Wang, K. K. W.; Narayana, S. V. L., A Structural Model for the Inhibition of Calpain by Calpastatin: Crystal Structures of the Native Domain VI of Calpain and its Complexes with Calpastatin Peptide and a Small Molecule Inhibitor. *J. Mol. Biol.* **2003,** *328*, 131-146.

12. Reverter, D.; Braun, M.; Fernandez-Catalan, C.; Strobl, S.; Sorimachi, H.; Bode, W., Flexibility Analysis and Structure Comparison of Two Crystal Forms of Calcium-Free Human m-Calpain. *Biol. Chem.* **2002,** *383*, 1415-1422.

13. Moldoveanu, T.; Hosfield, C. M.; Lim, D.; Elce, J. S.; Jia, Z.; Davies, P. L., A Ca2+ Switch Aligns the Active Site of Calpain. *Cell* **2002,** *108*, 649-660.

14. Wang, K. K. W.; Nath, R.; Posner, A.; Raser, K. J.; Buroker-Kilgore, M.; Hajimohammadreza, I.; Jr, A. W. P.; Marcoux, F. W.; Q. Ye, E. T.; Hatanaka, M.; Maki, M.; Caner, H.; Collins, J. L.; Fergus, A.; Lee, K. S.; Lunney, E. A.; Hays, S. J.; Yuen, P., An alpha-mercaptoacrylic acid derivative is a selective nonpeptide cell-permeable calpain inhibitor and is neuroprotective. *Proc. Natl. Acad. Sci. USA* **1996,** *93*, 6687-6692.

15. Inoue, J.; Nakamura, M.; Cui, Y.; Sakai, Y.; Sakai, O.; Hill, J. R.; Wang, K. K. W.; Yuen, P., Structure-Activity Relationship and Drug Profile of N-(4-Fluorophenylsulfonyl)- L valyl-L-leucinal (SJA6017) as a Potent Calpain Inhibitor. *J. Med. Chem.* **2003,** *46*, 868-871.

16. Hooper, N. M., Proteases: a primer. *Essays in Biochemistry* **2002,** *38*, 1-8.

17. Petrassi, H. M.; Williams, J. A.; Li, J.; Tumanut, C.; Ek, J.; Nakai, T.; Masick, B.; Backes, B. J.; Harris, J. L., A strategy to profile prime and non-prime proteolytic substrate specificity. *Bioorganic & Medicinal Chemistry Letters* **2005,** *15*, 3162–3166.

18. Moldoveanu, T.; Hosfield, C. M.; Lim, D.; Jia, Z.; Davies, P. L., Calpain Silencing by a Reversible Intrinsic Mechanism. *Nat. Struct. Biol.* **2003,** *10* 371-378.

19. Biswas, S.; Harris, F.; Dennison, S.; Singh, J.; Phoenix, D. A., Calpains: Targets of Cataract Prevention? *TRENDS in Molecular Medicine 2004* **2004,** *10*, 78-84.

20. Whittaker, M., Discovery of Protease Inhibitors Using Targeted Libraries. *Curr. Opin. Chem. Biol.* **1998,** *2*, 386-396.

21. Lilly, E.; Lilly, S., Stereoselective Synthesis of Aminoethylamine Aspartyl Protease Transition State Isosteres *Tetrahedron letters* **2006,** *47*, 7097-7100.

22. Lindvall, M. K., Molecular Modeling in Cysteine Protease Inhibitor Design. *Curr. Pharmaceutical Design* **2002,** *8*, 1673-1681.

23. Schneck, J. L.; Villa, J. P.; McDevitt, P.; McQueney, M. S.; Thrall, S. H.; Meek, T. D., Chemical Mechanism of a Cysteine Protease, Cathepsin C, As Revealed by Integration of both Steady-State and Pre-Steady-State Solvent Kinetic Isotope Effects. *Biochemistry* **2008,** *47*, 8697–8710.

24. Sims, A.; Dunn-Coleman, N.; Robson, G.; Oliver, S., Glutamic protease distribution is limited to filamentous fungi. *FEMS Microbiological Letters* **2004,** *239*, 95-101.

25. Flipo, M.; Florent, I.; Grellier, P.; Sergheraerta, C.; Deprez-Poulaina, R., Design, Synthesis and Antimalarial Activity of Novel, Quinoline-Based, Zinc Metallo-Aminopeptidase Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2003,** *13*, 2659–2662.

26. Flipo, M.; Beghyn, T.; Leroux, V.; Florent, I.; Deprez, B. P.; Deprez-Poulain, R. F., Novel Selective Inhibitors of the Zinc Plasmodial Aminopeptidase PfA-M1 as Potential Antimalarial Agents. *J. Med. Chem.* **2007,** *50*, 1322-1334.

27. Flipo, M.; Beghyn, T.; Charton, J.; Leroux, V. A.; Deprez, B. P.; Deprez-Poulain, R. F., A library of novel hydroxamic acids targeting the metallo-protease family: Design, parallel synthesis and screening. *Bioorganic & Medicinal Chemistry* **2007,** *15*, 63-76.

28. Johnson, S. L.; Chen, L.-H.; Pellecchia, M., A high-throughput screening approach to anthrax lethal factor inhibition. *Bioorganic Chemistry* **2007,** *35*, 306–312.

29. Klauss, V.; Spannagl, M., Thrombin Inhibitors and Anti-Factor Xa Agents in the Treatment of Arterial Occlusion. *Current Drug Targets* **2006,** *7*, 1285-1290.

30. Ingallinella, P.; Fattori, D.; Altamura, S.; Steinkühler, C.; Koch, U.; Cicero, D.; Bazzo, R.; Cortese, R.; Bianchi, E.; Pessi, A., Prime Site Binding Inhibitors of a Serine Protease: NS3/4A of Hepatitis C Virus. *American Chemical Society* **2002,** *41* 5483 -5492.

31. Kim, D.; Wang, L.; Beconi, M.; Eiermann, G. J.; Fisher, M. H.; He, H.; Hickey, G. J.; Kowalchick, J. E.; Leiting, B.; Lyons, K.; Marsilio, F.; McCann, M. E.; Patel, R. A.; Petrov, A.; Scapin, G.; Sangita B. Patel; Roy, R. S.; Wu, J. K.; Wyvratt, M. J.; Zhang, B. B.; Zhu, L.; Thornberry, N. A.; Weber, A. E., (2R)-4-Oxo-4-[3-(Trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl]-1-(2,4,5-trifluorophenyl)butan-2-amine: A Potent, Orally Active Dipeptidyl Peptidase IV Inhibitor for the Treatment of Type 2 Diabetes. *J. Med. Chem.* **2005,** *48*, 141-151.

32. Zavrski, I.; Jakob, C.; Schmid, P.; Krebbel, H.; Kaiser, M.; Fleissner, C.; Rosche, M.; Possinger, K.; Sezer, O., Proteasome: an emerging target for cancer therapy. *Anti-Cancer Drugs* **2005,** *16*, 475-481.

33. Barrett, A. J.; Rawlings, N. D., Evolutionary Lines of Cysteine Peptidases. *Biol. Chem.* **2001,** *382*, 727 – 733.

34. Donker, I. O., A Survey of Calpain Inhibitors. *Curr. Med. Chem.* **2000,** *7* 1171-1188.

35. Iqbal, M.; Messina, P. A.; Freed, B.; Das, M.; Chatterjee, S.; Tripathy, R.; Tao, M.; Josef, K. A.; Dembofsky, B., Subsite requirements for peptide aldehyde inhibitors of human calpain I. *Bioorganic & Medicinal Chemistry Letters* **1997,** *7*, 539-544.

36. Perrin, B. J.; Huttenlocher, A., Calpain. *Int. J. Biochem Cell Biol.* **2002,** *34*, 722–725.

37. Arad, D.; Langridge, R.; Kollman, P. A., A Simulation of the Sulfur Attack in the Catalytic Pathway of Papain Using Molecular Mechanics and Semiempirical Quantum Mechanics. *J. Am. Chem. Soc.* **1990,** *112*, 491-502.

38. Reid, R. C.; Pattenden, L. K.; Tyndall, J. D. A.; Martin, J. L.; Walsh, T.; Fairlie, D. P., Countering Cooperative Effects in Protease Inhibitors Using Constrained ß-Strand-Mimicking Templates in Focused Combinatorial Libraries. *J. Med. Chem.* **2004,** *47* 1641-1651.

39. Glenn, P.; Pattenden, L. K.; Reid, R. C.; Tyssen, D. P.; Tyndall, J. D. A.; Birch, C. J.; Fairlie, D. P., ß-Strand Mimicking Macrocyclic Amino Acids: Templates for Protease Inhibitors with Antiviral Activity. *J. Med. Chem.* **2002,** *45* 371-381.

40. Fairlie, D. P.; Tyndall, J. D. A.; Reid, R. C.; Wong, A. K.; Abbenante, G.; Scanlon, M. J.; March, D. R.; Bergman, D. A.; Chai, C. L. L.; Burkett, B. A., Conformational Selection of Inhibitors and Substrates by Proteolytic Enzymes: Implications for Drug Design and Polypeptide Processing. *J. Med. Chem.* **2000,** *43*, 1271-1281.

41. Lucke, A. J.; Tyndall, J. D. A.; Fairlie, D. P., Designing Supramolecular Structures from models of Cyclic Peptide Scaffolds with Heterocyclic Constraints. *Journal of Molecular Graphics and Modelling* **2003,** *21*, 341-355.

42. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyaov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000,** *28* 235-242.

43. Moldoveanu, T.; Campbell, R. L.; Cuerrier, D.; Davies, P. L., Crystal Structures of Calpain-E64 and-Leupeptin Inhibitor Complexes Reveal Mobile Loops Gating the Active Site. *J. Mol. Biol.* **2004,** *343* 1313.

44. *MacroModel*, version 9.1; Schrödinger: LLC, New York, NY, 2005.

45. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energies and Properties of Organic Liquids. *JACS* **1996,** *118* 11225.

46. Kolossváry, I.; Guida, W. C., Low-Mode Conformational Search Elucidated: Application to C39H80 and Flexible Docking of 9-Deazaguanine Inhibitors into PNP. *J. Comp. Chem.* **1999,** *20* 1671-1684.

47. Kolossváry, I.; Guida, W. C., Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and cyclic Peptides. *JACS* **1996,** *118*, 5011-5019.

48. Qui, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A.* **1997,** *101* 3005.

49. Bostrom, J., Reproducing the Conformations of Protein-Bound Ligands: A Critical Evaluation of Several Popular Conformational Searching Tools. *J. Comput. Aid. Mol. Des.* **2001,** *15* 1137-1152.

50. Shenkin, P. S.; McDonald, D. Q., Cluster Analysis of Molecular Conformations. *J. Comp. Chem.* **2004,** *15* 899-916.

51. *MacroModel XCluster*, version 9.1; Schrödinger: LLC, New York, NY, 2005.

52. Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D., Comparative Evaluation of Eight Docking and Virtual Screening Accuracy. *PROTEINS: Structure, Function, and Bioinformatics* **2004,** *57*, 225-242.

53. Kontoyianni, M.; McClellan, L. M.; Sokol, G. S., Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* **2004,** *47* 558-565.

54. Schulz-Gasch, T.; Stahl, M., Binding Site Characteristics in Structure-based Virtual Screening: Evaluation of Current Docking Tools. *J. Mol. Model.* **2003,** *9*, 47-57.

55. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004,** *47*, 1750-1759.

56. Dainese, E.; Minafra, R.; Sabatucci, A.; Vachette, P.; Melloni, E.; Cozzani, I., Conformational Changes of Calpain from Human Erythrocytes in the Presence of Ca2+. *J. Biol. Chem.* **2002,** *277*, 40296-40301.

57. *Schrödinger Suite 2006 Induced Fit Docking Protocol*, Glide version 4.0, Prime version 1.5; Schrödinger: LLC, New York, NY, 2005.

58. Cavasotto, C. N.; Abagyan, R. A., Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004,** *337* 209-225.

59. *Glide*, version 4.0; Schrödinger: LLC, New York, NY, 2005.

60. *Prime*, version 1.5; Schrödinger: LLC, New York, NY, 2005.

2 Development of a calpain model for docking studies

2.1 Introduction

2.1.1 Overview of calpain model development

My work has involved the use of the modeling programs MacroModel (molecular mechanics),[1-5] Xcluster (cluster analysis),[6, 7] Glide (automated docking),[8, 9] Prime (protein structure prediction, preparation and refinement),[10] and the InducedFit Docking Protocol (a combination of Glide and Prime to induce protein flexibility while docking ligands).[11, 12]

A methodology has been devised that can be applied specifically to the calpain/cataract problem:

(i) The calpain enzymes and other cysteine proteases in the Protein Data Bank (PDB)[13] have been identified, and the active sites in each examined in detail. All the calpain structures that were available at the start of my studies were found to be in a non-active conformation, however, several other cysteine proteases, such as papain, have the catalytic triad in the active conformation.

In the active conformation the cysteine proteases facilitate peptide hydrolysis of specific sites in a protein target. Soon after the project commenced , the structural data of two calpain constructs (PDB code 1KXR and 1MDW) that display an active catalytic triad conformation, were deposited in the PDB.[14, 15] This was timely for our research. These two constructs consist of only domains I and II of μ-calpain and m-calpain,

respectively. IKXR was chosen for further study because this construct displayed activity *in vitro* whereas 1MDW did not.

(ii) The literature was examined to establish what was known about the mechanism of cysteine protease (calpain) peptide hydrolysis and in particular the role of the catalytic triad of the active enzymes.[16-18] This mechanistic information has allowed me to use the X-ray structural data and prepare a model of the active site of a calpain enzyme. Quantum mechanical studies of the mechanism of cysteine protease hydrolysis has also been used to prepare the model for subsequent docking. The mechanism of action of cysteine protease being one where the active site thiol is first deprotonated and the active site histidine protonated upon the binding of a substrate.[16-18]

2.2 Calpain X-ray crystal structures

2.2.1 The first published structures

The first X-ray crystal structures of calpain (1AJ5 and 1DVI) were deposited in the PDB[13] by Blanchard *et al*[19] in the late 1990's. The structures were not of the complete enzyme but of the Ca^{2+}-binding domain VI of rat m-calpain in the absence of Ca^{2+} (PDB code 1AJ5) and with Ca^{2+} bound (1DVI). The X-ray structures revealed that domain VI contained five EF-hand motifs and that three of them are bound to Ca^{2+}. EF-hands are helix-loop-helix structural motifs which are known to commonly bind Ca^{2+}. The presence of structural changes induced by Ca^{2+} binding at the EF-hands gave critical insights into how Ca^{2+} binding causes activation of calpain. As already mentioned it is now known that Ca^{2+} binding to

domain VI (**Figure 2.1**) and the structurally similar domain IV, induces structural changes in the active site which are necessary for activation of the enzyme.

Lin *et al*[20] went one step further and solved the structure of domain VI of porcine m-calpain with Ca^{2+} bound (1ALV) and with Ca^{2+} and an inhibitor PD150606 bound (1ALW). The inhibitor, PD150606 (3-(4-iodophenyl)-2-mercapto-(Z)-2-propenoic acid) (**Figure 2.2**) was earlier discovered and proposed to be a non-active site inhibitor by Wang *et al*.[21] It binds into a hydrophobic pocket created by three helices (right **Figure 2.1**) and this pocket was later discovered to be where residue Phe610 of calpastatin (the endogenous calpain inhibitor) also binds.[22]

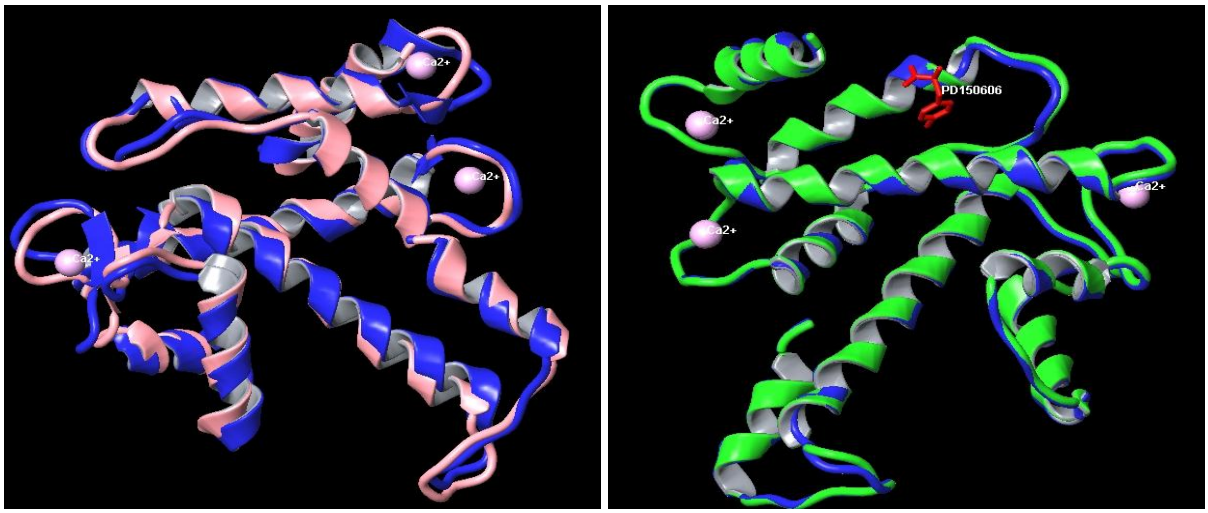


Figure 2.1: Left – Cartoon ribbon diagram of domain VI of rat m-calpain in the absence of Ca^{2+} (blue ribbon) superimposed on domain VI of rat m-calpain with Ca^{2+} bound (pink ribbon) (Ca^{2+} = pink spheres). Right – Cartoon ribbon diagram of domain VI of porcine m-calpain with Ca^{2+} bound (green ribbon) superimposed on domain VI of porcine m-calpain with Ca^{2+} and an inhibitor bound (blue ribbon) (Ca^{2+} = pink spheres, PD150606 = red tube).

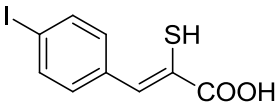


Figure 2.2: Structure of PD150606

The first crystal structure of an entire calpain heterodimer was reported by Hosfield *et al* in 1999.[23] The structure was of rat m-calpain in the absence of Ca^{2+} (1DFO) and confirmed for the first time that the catalytic triad was not in an active conformation when Ca^{2+} was absent. Ca^{2+} is required to be present for activation to occur. This structure (1DFO) was soon followed by the publication of two structures of human m-calpain (1KFU and 1KFX) both in the absence of Ca^{2+} and these confirmed the existence of a disrupted active site for the apo-enzyme.[24]

There are severe problems associated with co-crystallising native μ-calpain and m-calpain in the presence of Ca^{2+} because they are autoproteolytic (autocatalytic enzymes), meaning they literally chop themselves into fragments which renders them inactive.[14, 15, 25] This autocatalysis is part of the enzymes natural cycle of activation and regulation by factors such as Ca^{2+}, calpastatin, the interaction with lipids/membranes, and activator proteins. Autocatalysis therefore makes it impossible to obtain a crystal structure of native heterodimeric calpain with Ca^{2+} bound and, hence, in an active conformation. This fact also makes it impossible to obtain a structure with an inhibitor bound into the active site of native heterodimeric calpain.

2.2.2 Calpain constructs 1KXR and 1MDW

A breakthrough was made in 2002 when Moldoveanu *et al*[14] crystallised the protease core of μ-calpain (1KXR) with the active site Cys115 mutated to Ser115 in the presence of Ca^{2+}. The construct was mutated by Moldoveanu *et al*[14] from Cys115 to Ser115 to inactivate

the construct (prevent autolysis) during crystallization. The protease core of this mutated μ-calpain consists of domains I and II which contain the active site of the enzyme.

The non-mutated construct, analogous to 1KXR, is generated during autolysis as a stable calpain fragment. Moldoveanu showed that this non-mutated μ-I-II construct is inactive in the absence of Ca^{2+} but active in the presence of Ca^{2+} when tested *in vitro*.

Examination of the active site of the mutated construct (1KXR) showed that the catalytic triad was in a conformation analogous to that seen in other non calcium activated cysteine proteases, including papain.

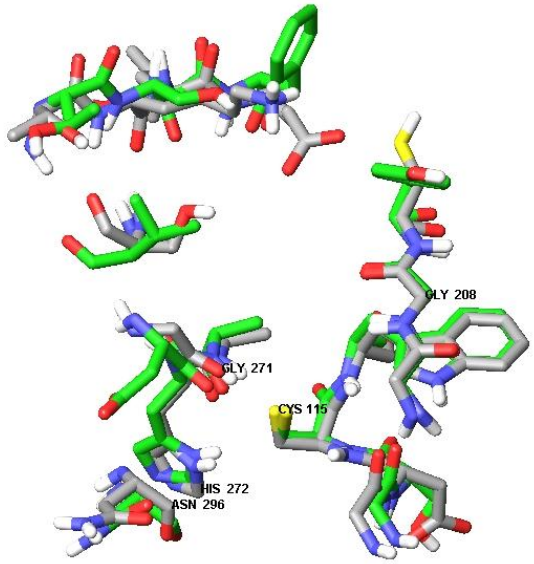


Figure 2.3: The active site of μ-calpain construct (1KXR) with grey carbons, superimposed on active site of papain (1PPN) with green carbons. Residues named are for 1KXR, including the catalytic triad of Cys115 (mutated *in silico* from Ser115), His272, and Asn296. The equivalent papain residues are Cys25, His272, His159, and Asn175, respectively.

The active site of the μ-calpain construct (1KXR) with cysteine 115 mutated to serine 115 superimposed on the active site of papain (1PPN) is shown in **Figure 2.3**. The two active sites show striking similarity, emphasized by the close superimposition of critical residues

including the catalytic triad of Cys115 (papain Cys25), His272 (papain His159), and Asn296 (papain Asn175).

For our studies the crystal structure (1KXR) was mutated *in silico* with the Ser115 replaced by Cys115 to produce a model that approximates that of the active construct.

Moldoveanu *et al* also crystallized the protease core of m-calpain (1MDW) in the presence of Ca^{2+}.[15] However, the m-calpain construct was not active in the presence of Ca^{2+} *in vitro*. The crystal structure revealed that an intrinsic mechanism causes the autolysis-generated protease core fragment of m-calpain to be inactivated through inherent instability of a key α-helix.[15] Residues 198-201 within this helical region (α7) display no electron density in 1MDW which is usually seen in native m-calpain structures and in the μ-calpain construct (1KXR) (**Figure 2.4**).

Moldoveanu *et al*[15] proposed that the flexibility of this glycine rich region (residues 197-203) permits the side chain of Trp106 to swing around and into the active site pocket, as is seen in **Figure 2.5,** giving a structural basis for auto-inhibition. This is not the case with other isoforms of calpain such as μ-calpain.

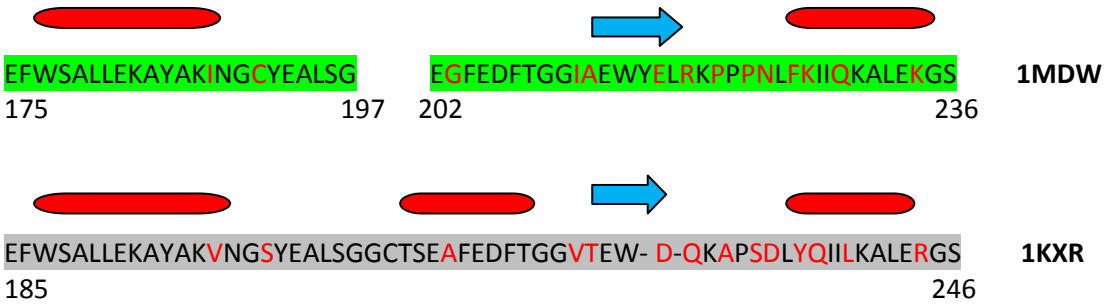


Figure 2.4: Part sequence of 1MDW (residues 175-236) and the equivalent part sequence of 1KXR (residues 185-246). The red tubes indicate regions of α-helices, blue arrows indicate regions of β-sheets. The red residue letter codes indicate non-conserved residues between μ-calpain and m-calpain. The region in 1MDW between residues 197 and 202 is normally occupied by the sequence GATT in m-calpain and is missing in 1MDW as there was no visible electron density.[14, 15]

At the beginning of this project the μ-calpain construct (1KXR) was the only X-ray crystal structure of calpain in the active conformation. This structure was therefore used as a template to develop a model for our initial inhibitor docking studies. At the same time, papain was also examined in detail due to the similarity of the active site with that in calpain. There are many X-ray crystal structures of papain in the PDB with and without cysteine protease inhibitors bound in the active site and in the β-strand conformation. Many of these inhibitors have been shown to also inhibit both μ-calpain and m-calpain.

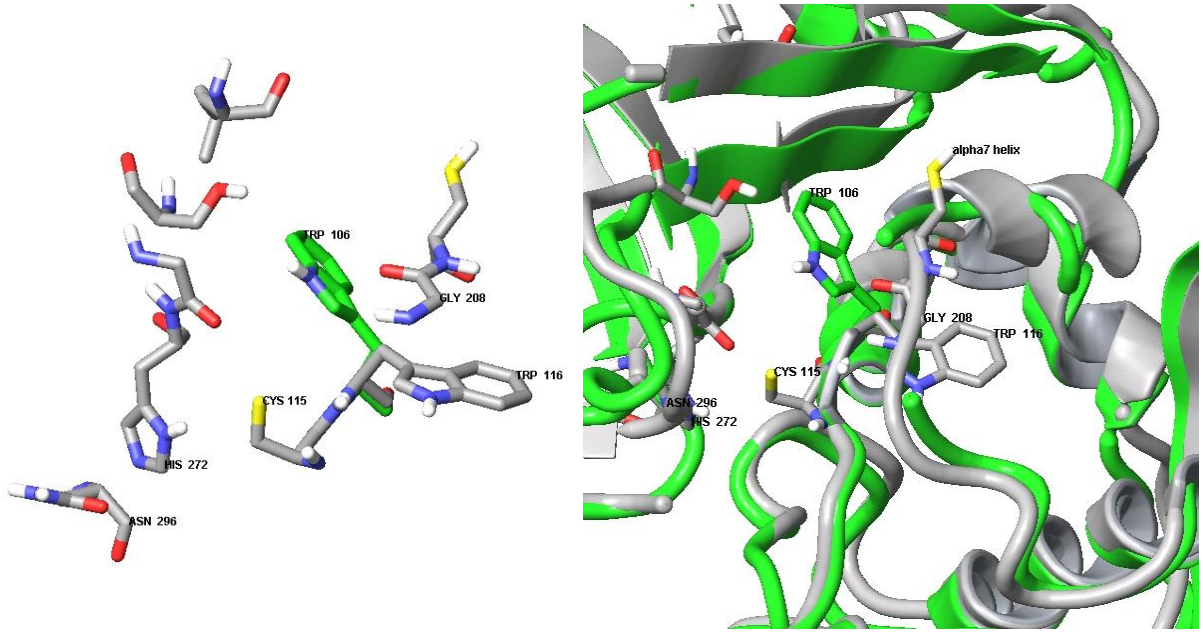


Figure 2.5: left – active site of 1KXR (grey) superimposed on active site of 1MDW (green) showing Trp116 of 1KXR in the active conformation and the equivalent Trp106 of 1MDW blocking the active site. Right - active site of 1KXR (grey) superimposed on active site of 1MDW (green) showing the α7 helix of 1KXR intact and the equivalent region of 1MDW in a disordered state with missing residues due to no observable electron density.

2.3 Exploring the calpain construct 1KXR to develop a viable model for Glide docking experiments

The X-ray crystal structure 1KXR, the first structure of calpain with the active site in an active conformation was the only such structure available early in this project. The active site of 1KXR was explored using Schrodinger's molecular modeling suite.

The coordinates of 1KXR were downloaded from the PDB and imported into Maestro[26], Schrodinger's graphical interface. The structure was 'cleaned up' by deleting all the water molecules (this is generally performed as solvation effects are considered in the empirical scoring function, however, specific waters may be included when they form a bridge between ligand and protein) and deleting the second molecule of the dimer (chain B) to leave the monomer (chain A) from the crystal structure. Hydrogens were added to this chain A and the active site Ser115 mutated *in silico* to Cys115 in order to produce a structure that resembles the active site of the native enzyme. As shown from *ab initio* studies the active site thiol is thought to be deprotonated and the active site histidine protonated upon the initial binding of a substrate.[16-18] Deprotonation of Cys115 and protonation of His272 was performed *in silico* to simulate this mechanistic change upon ligand binding. The rationale here is that the deprotonated cysteine model would approximate the active site at the point immediately before nucleophilic attack. A full description of the model refinement is described in **Protocol 1** in the **Appendix**. The output structure of this minimisation was subsequently used for docking experiments and is referred to in this thesis as the Glide model.

The Glide model was examined in detail before docking experiments were performed to familiarise myself with the enzymes features, especially the active site and surrounding residues. **Figure 2.6** shows the active site of the Glide model. It is a narrow cleft bordered by two steep sides which are made up of domain I (left side of each picture) and domain II (right side of each picture). The two ends of the cleft, with Cys115 being the starting point, are known as the prime side and the non-prime side. The non-prime side is where most of the known inhibitors of calpain bind the enzyme.

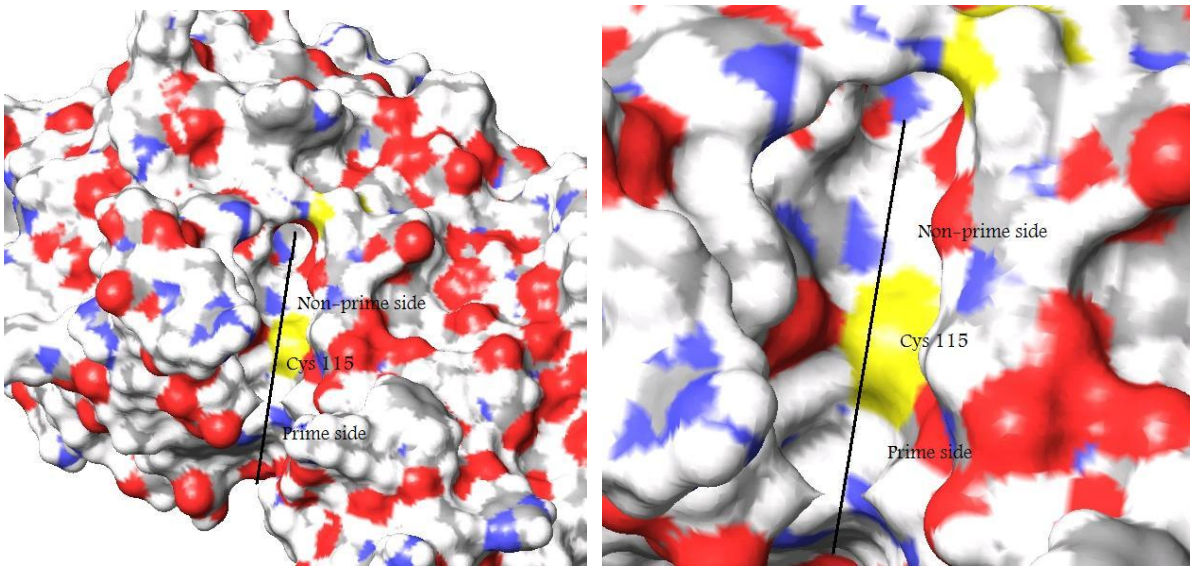


Figure 2.6: Left – Surface diagram of 1KXR (Glide model) showing the active site cleft (black line) with the active site Cys 115 in yellow. Above Cys 115 is the non-prime side and below is the prime side. Right – A close up diagram of the top left diagram. The active site is a deep valley with high sides.

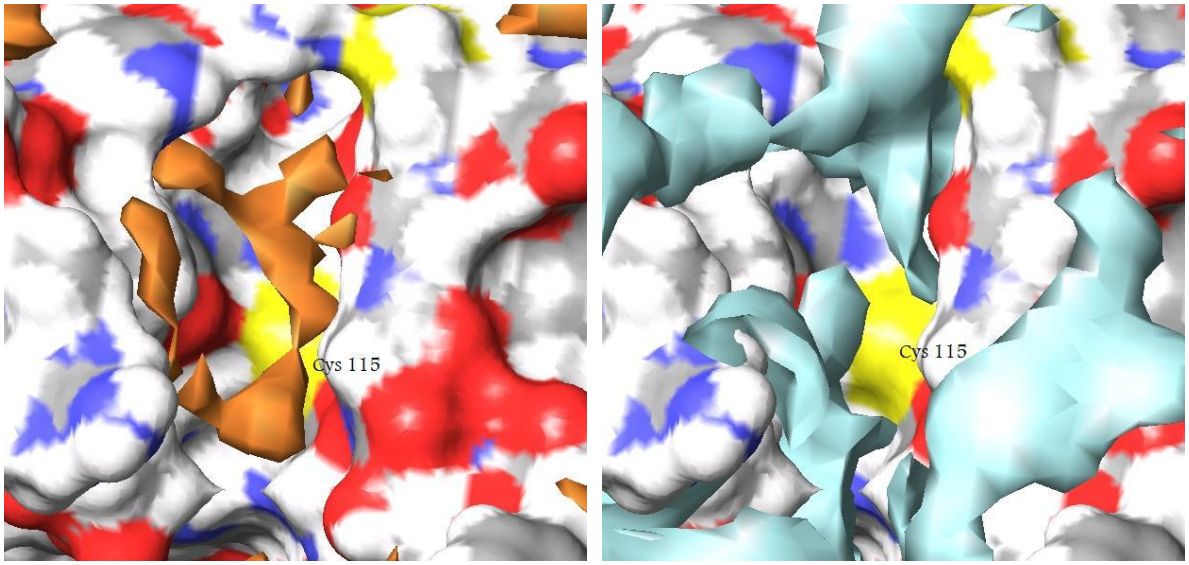


Figure 2.7: Left – Copper areas depict regions of space on the surface of the enzyme that are hydrophobic. Deep within the non-prime side are areas of hydrophobicity. Right – Pale blue areas depict regions of space on the surface of the enzyme that are hydrophilic. Some surrounding areas just out of the pocket are hydrophilic.

Deep within the non-prime region are areas of hydrophobicity which can be seen in the left picture of **Figure 2.7** depicted as copper coloured regions. These regions are likely to accommodate hydrophobic amino acid side chains of a bound ligand.

Areas of hydrophilicity can be seen in the right picture of **Figure 2.7** as pale blue areas surrounding the active site. These areas could accommodate hydrophilic groups of an inhibitor when bound into the active site.

The amino acid residues that surround the non-prime region of the active site are shown in **Figure 2.8**. Cys115 is the active site nucleophilic residue that facilitates cleavage of peptide substrates. On either side of Cys115 are the residues Gly208 and Gly271 which are now known to form H-bonds to peptide based ligands that form the β-strand conformation but were suspected to be involved earlier based on co-crystallised structures of other cysteine proteases such as papain with inhibitors bound. The three H-bonds formed between the ligand and these two residues are thought to be essential for stabilising the bound ligand.

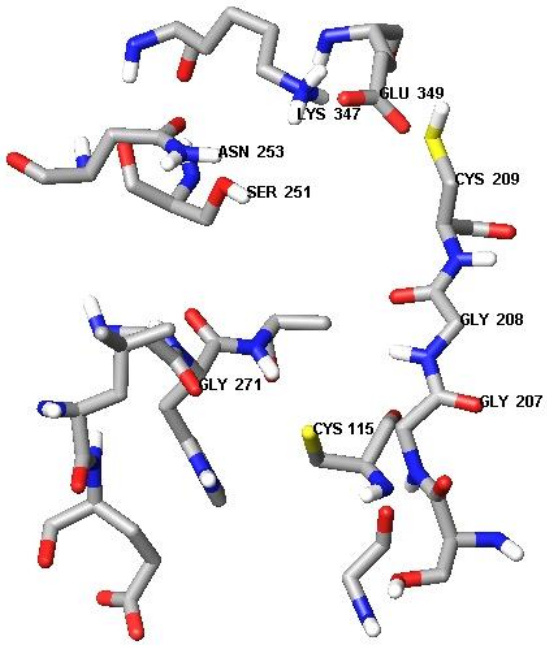


Figure 2.8: Active site of the calpain Glide model based on the PDB structure 1KXR showing the residues surrounding the non-prime region.

Di- and tripeptide based inhibitors of calpain usually have some type of capping group at the N-terminus which usually positions itself around the S_3 subsite of the enzyme. This P_3

subsite is surrounded by the residues Cys209, Ser251, Asn253, Lys347, and Glu349 all of which have H-bond donators or acceptors, therefore, capping groups which have H-bond acceptors or donators would likely be able to form H-bonds with these residues. Such H-bond formation would be favourable and likely increase the binding affinity of the ligand.

2.4 The InducedFit docking model

The Glide model had to be slightly modified to work with the InducedFit Protocol because Prime (part of the InducedFit protocol) would not allow the deprotonated Cys115 to work in the docking procedure. Therefore, Cys115 was left protonated in the model used for InducedFit docking and is referred to in this thesis as the InducedFit model. The refinement of the InducedFit model is described in **Protocol 2** in the **Appendix**.

2.5 Conclusion

X-ray crystal structures of calpains and other cysteine proteases such as papain were examined in detail to gain knowledge of the calpain enzymes active site. Calpain structures with inhibitors bound into the active site were not available at the beginning of this project. However, papain, a structurally similar cysteine protease, co-crystallised with inhibitors bound gave insight into the way peptidomimetic inhibitors bind the active site, especially how they bind in a β-strand conformation and with three essential H-bonds between the ligand backbone and two glycine residues of the enzyme.

It was also important to learn that most of the inhibitors that bind the active site of papain bind to the non-prime side and so it was reasonable to assume these inhibitors and their

analogues would bind calpain on this side. This information was critical for setting up the docking experiments with the aim of designing inhibitors as a docking grid of the active site must be produced before ligands are docked. The docking grid is a set of coordinates that isolates and maps the site on the enzyme to where ligands are to be docked. From these structures I was also able to propose modifications to known inhibitors based on the available space and environment around the bound inhibitors and active site.

The timely release of the 1KXR structure was essential to the development of a working model for use in subsequent docking experiments with Glide and the InducedFit Protocol. This was because 1KXR was the first X-ray crystal structure of a calpain enzyme in the active conformation which meant that ligands could be docked in order to evaluate their potential as calpain inhibitors.

Several structures of the calpain construct with inhibitors bound in the active site have recently been published. These structures have an advantage over 1KXR because their active sites have made subtle changes to accommodate their bound ligands and, therefore, would make for good models for docking experiments of similar ligands. Future work to develop and test such models would be an essential step to developing the next generation of calpain inhibitors.

2.6 References

1. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energies and Properties of Organic Liquids. *JACS* **1996,** *118* 11225.
2. Kolossváry, I.; Guida, W. C., Low-Mode Conformational Search Elucidated: Application to C39H80 and Flexible Docking of 9-Deazaguanine Inhibitors into PNP. *J. Comp. Chem.* **1999,** *20* 1671-1684.
3. Kolossváry, I.; Guida, W. C., Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and cyclic Peptides. *JACS* **1996,** *118*, 5011-5019.
4. Qui, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A.* **1997,** *101* 3005.
5. *MacroModel*, version 9.1; Schrödinger: LLC, New York, NY, 2005.
6. Shenkin, P. S.; McDonald, D. Q., Cluster Analysis of Molecular Conformations. *J. Comp. Chem.* **2004,** *15* 899-916.
7. *MacroModel XCluster*, version 9.1; Schrödinger: LLC, New York, NY, 2005.
8. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004,** *47*, 1750-1759.
9. *Glide*, version 4.0; Schrödinger: LLC, New York, NY, 2005.
10. *Prime*, version 1.5; Schrödinger: LLC, New York, NY, 2005.

11. *Schrödinger Suite 2006 Induced Fit Docking Protocol*, Glide version 4.0, Prime version 1.5; Schrödinger: LLC, New York, NY, 2005.

12. Cavasotto, C. N.; Abagyan, R. A., Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004,** *337* 209-225.

13. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyaov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000,** *28* 235-242.

14. Moldoveanu, T.; Hosfield, C. M.; Lim, D.; Elce, J. S.; Jia, Z.; Davies, P. L., A Ca2+ Switch Aligns the Active Site of Calpain. *Cell* **2002,** *108*, 649-660.

15. Moldoveanu, T.; Hosfield, C. M.; Lim, D.; Jia, Z.; Davies, P. L., Calpain Silencing by a Reversible Intrinsic Mechanism. *Nat. Struct. Biol.* **2003,** *10* 371-378.

16. Arad, D.; Langridge, R.; Kollman, P. A., A Simulation of the Sulfur Attack in the Catalytic Pathway of Papain Using Molecular Mechanics and Semiempirical Quantum Mechanics. *J. Am. Chem. Soc.* **1990,** *112*, 491-502.

17. Nemukhin, A. V.; Grigorenko, B. L.; Rogov, A. V.; Topol, I. A.; Burt, S. K., Modeling of Serine Protease Prototype Reactions with the Flexible Effective Fragment Potential Quantum Mechanical/Molecular Mechanical Method. *Theor. Chem. Acc.* **2004,** *111* 36-48.

18. Topf, M.; Varnai, P.; Richards, W. G., Ab Initio QM/MM Dynamics Simulation of the Tetrahedral Intermediate of Serine Proteases: Insights into the Active Site Hydrogen-Bonding Network. *J. Am. Chem. Soc.* **2002,** *124* 14780-14788.

19. Blanchard, H.; Grochulski, P.; Li, Y.; Arthur, J. S. C.; Davies, P. L.; Elce, J. S.; Cygler, M., Structure of a calpain Ca2+-binding domain reveals a novel EF-hand and Ca2+-induced conformational changes. *Nature Structural Biology* **1997,** *4*, 532 - 538.

20. Lin, G.-d.; Chattopadhyay, D.; Maki, M.; Wang, K. K. W.; Carson, M.; Jin, L.; Yuen, P.-w.; Takano, E.; Hatanaka, M.; DeLucas, L. J.; Narayana, S. V. L., Crystal structure of calcium bound domain VI of calpain at 1.9 Å resolution and its role in enzyme assembly, regulation, and inhibitor binding. *Nature Structural Biology* **1997,** *4*, 539 - 547.

21. Wang, K. K. W.; Nath, R.; Posner, A.; Raser, K. J.; Buroker-Kilgore, M.; Hajimohammadreza, I.; Jr, A. W. P.; Marcoux, F. W.; Q. Ye, E. T.; Hatanaka, M.; Maki, M.; Caner, H.; Collins, J. L.; Fergus, A.; Lee, K. S.; Lunney, E. A.; Hays, S. J.; Yuen, P., An alpha-mercaptoacrylic acid derivative is a selective nonpeptide cell-permeable calpain inhibitor and is neuroprotective. *Proc. Natl. Acad. Sci. USA* **1996,** *93*, 6687-6692.

22. Todd, B.; Moore, D.; Deivanayagam, C. C. S.; Lin, G.; Chattopadhyay, D.; Maki, M.; Wang, K. K. W.; Narayana, S. V. L., A Structural Model for the Inhibition of Calpain by Calpastatin: Crystal Structures of the Native Domain VI of Calpain and its Complexes with Calpastatin Peptide and a Small Molecule Inhibitor. *J. Mol. Biol.* **2003,** *328*, 131-146.

23. Hosfield, C. M.; Elce, J. S.; Davies, P. L.; Jia, Z., Crystal structure of calpain reveals the structural basis for Ca2+-dependent protease activity and a novel mode of enzyme activation. *EMBO* **1999,** *18*, 6880 - 6889.

24. Strobl, S.; Fernandez-Catalan, C.; Braun, M.; Huber, R.; Masumoto, H.; Nakagawa, K.; Irie, A.; Sorimachi, H.; Bourenkowi, G.; Bartuniki, H.; Suzuki§, K.; Bode, W., The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *PNAS* **2000,** *97*, 588–592.

25. Gabrijelcic-Geiger, D.; Mentele, R.; Meisel, B.; Hinz, H.; Assfalg-Machleidt, I.; Machleidt, W.; Möller, A.; Auerswald1, E. A., Human μ-Calpain: Simple Isolation

from Erythrocytes and Characterization of Autolysis Fragments. *Biol. Chem.* **2001,** *382*, 1733 – 1737.

26. *Maestro*, version 7.5; Schrödinger: LLC, New York, NY, 2005.

27. Moldoveanu, T.; Campbell, R. L.; Cuerrier, D.; Davies, P. L., Crystal Structures of Calpain-E64 and-Leupeptin Inhibitor Complexes Reveal Mobile Loops Gating the Active Site. *J. Mol. Biol.* **2004,** *343* 1313.

3 Molecular modeling of acyclic inhibitors

3.1 Introduction

3.1.1 Natural inhibitors

Among the naturally occurring calpain inhibitors is the endogenous polypeptide calpastatin which is found in mammalian cells along with μ-calpain and m-calpain.[1, 2] Calpastatin is composed of four repeating domains, which afford it the inhibitory qualities observed, and an N-terminal domain (domain L) that by itself displays no inhibition (**Figure 3.1**).[1] Within domains 1-4 are three highly conserved regions (regions A, B, and C). Region B can inhibit calpain on its own but is much more potent when flanked by regions A and C which are responsible for calcium dependant anchoring to calpain.[3, 4]

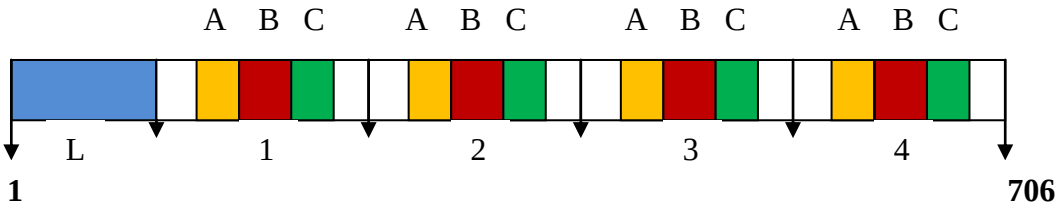
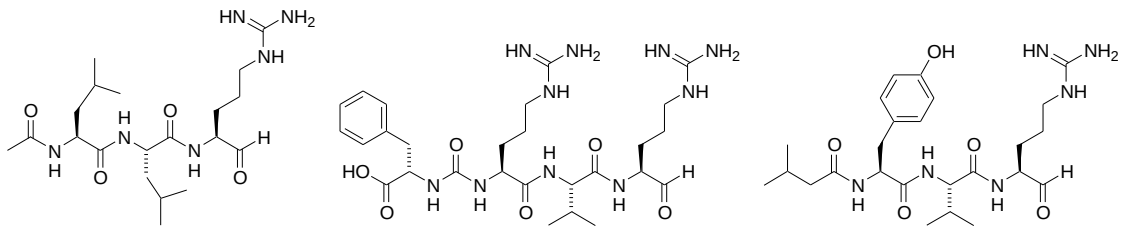


Figure 3.1: Structure of calpastatin.

There are also a number of calpain inhibitors derived from microbes. Many have been isolated from species of *Streptomyces* fungi including leupeptin (compound **3.1**), antipain (**3.2**), and strepin P-1 (**3.3**).[1] They are examples of calpain inhibitors that react reversibly with the active site thiol. Inhibitors of this type proved to be inadequate as biomedical tools to elucidate calpain cell function because they are non-selective and have poor membrane permeability,[1] characteristics that also render them poor drug candidates.



3.1 **3.2** **3.2**

Figure 3.2: Structures of leupeptin (compound **3.1**), antipain (**3.2**), and strepin P-1 (**3.3**), respectively.

However, the X-ray crystal structures of compounds such as leupeptin in complex with calpain have provided insight as to how they bind in or affect the active site of the enzyme. This has proved to be invaluable for the design of inhibitors. The X-ray crystal structure of the μ-calpain construct in complex with leupeptin (protein database 1TL9)[5] is shown in **Figure 3.3**. The left diagram shows the inhibitor with a covalent bond between the carbon of the aldehyde warhead and the sulphur of Cys115. There are three hydrogen bonds between the leupeptin backbone and Gly271 and Gly208 of the enzyme (bonds not shown). These three hydrogen bonds have been found in a number of inhibitor-calpain complexes, including the E64 - μ-calpain construct (1TL0)[5], SNJ-1715 - μ-calpain construct (2G8E)[6], and ZLAK-3001 - μ-calpain construct (2R9C)[7], alluding to their intrinsic necessity for potent inhibition.

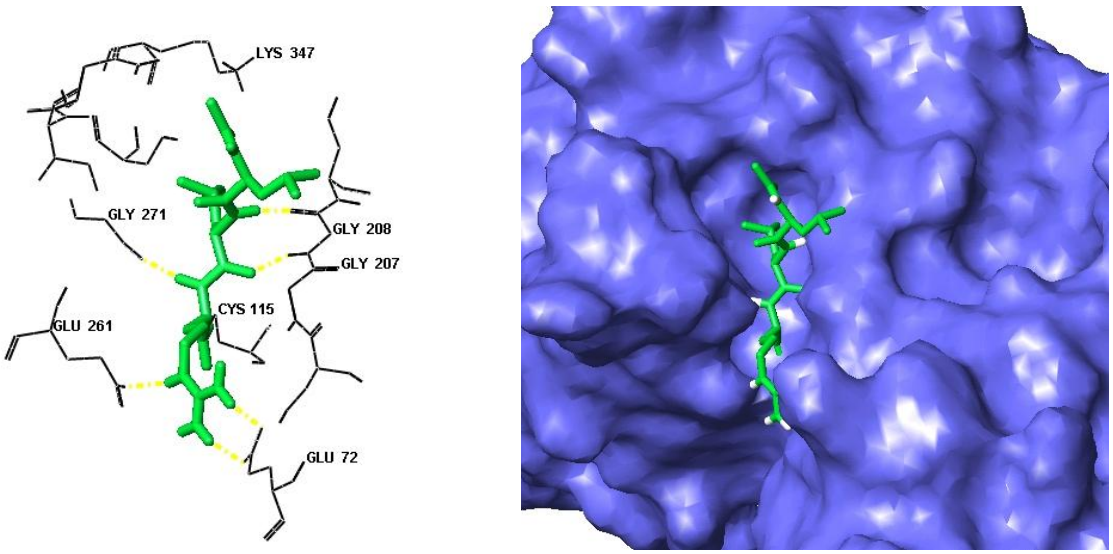
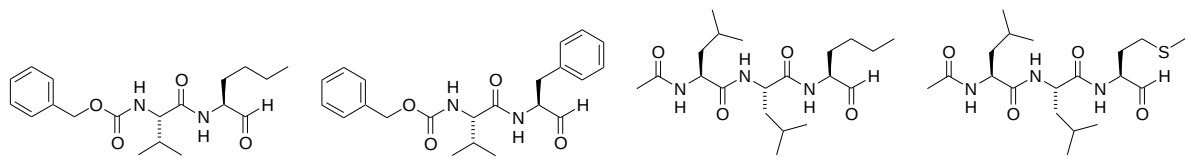


Figure 3.3: X-ray structure of Leupeptin (**3.1**) in complex with μ-calpain construct (1TL9). Left – enzyme depicted as sticks, leupeptin as tubes. Right – Surface model of enzyme with leupeptin as tubes.

The right hand diagram in **Figure 3.3** shows the surface of the enzyme bound with leupeptin. The inhibitor is stretched out along the cleft with its steep sides formed by domain I (righthand side) and domain II (lefthand side). It is worth noting that all of the co-crystallised X-ray structures of the μ-calpain construct in complex with bound inhibitors were first reported well into my thesis work but were invaluable in confirming the earlier hypothesis of a β-strand and also the assumption that three H-bonds to Gly208 and Gly271 were essential for tight ligand binding.

3.1.2 Modified natural inhibitors

To combat the lack of selectivity and permeability of the many natural inhibitors researchers first modified them by N-terminal capping with lipophilic substituents.[1] This led to several compounds exhibiting cell permeability including calpeptin (**3.4**),[8] MDL 28 (**3.5**),[9] calpain inhibitor I (**3.6**), and calpain inhibitor II (**3.7**) (**Figure 3.4**).[10] SAR (structure activity relationship) data showed these compounds to be more potent and cell permeable than the unmodified natural inhibitors (**Figure 3.2**) but they lacked selectivity for calpain over other proteases: for example, calpain inhibitor I is more potent against cathepsin than calpain. The aldehyde "warheads" are all too readily oxidized under cellular conditions.[1]



3.4 **3.5** **3.6** **3.7**

Figure 3.4: Structures of calpeptin (**3.4**), MDL 28 (**3.5**), calpain inhibitor I (**3.6**), and calpain inhibitor II (**3.7**).

3.1.3 Lead compound: SJA-6017

At the beginning of this project the compound SJA-6017[11] was used by our research group as a lead compound or starting structure from which to develop calpain inhibitors that could be more potent or specific *in vitro*. As well as SJA-6017, the most potent inhibitor of calpain known at the time, other known inhibitors were considered to identify features that make for good calpain inhibition. SJA-6017 and many of the known inhibitors contain three regions; a warhead at one end, a middle usually consisting of two or three amino acids acting as a *β*-strand, and an address region at the opposite end to the warhead.

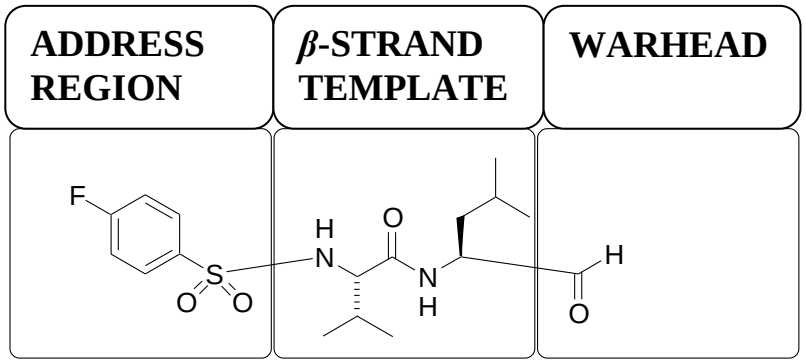


Figure 3.5: Diagram of the three main regions of a typical calpain inhibitor. Inhibitor shown is SJA-6017.

3.2 Docking studies of known inhibitors

3.2.1 Compounds of Inoue *et al*

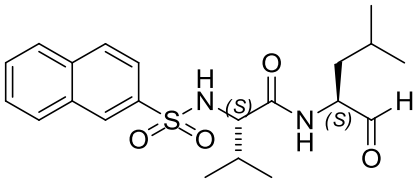
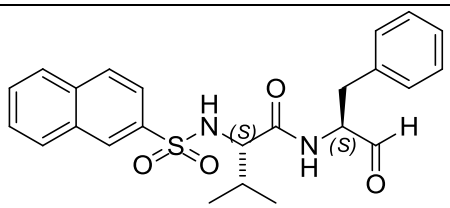
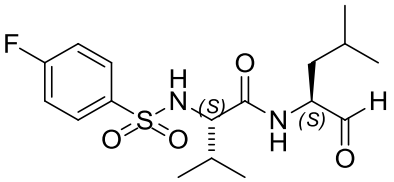
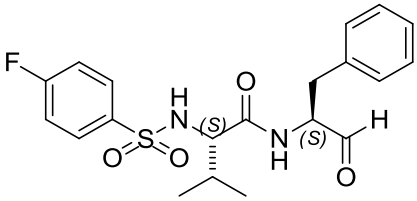
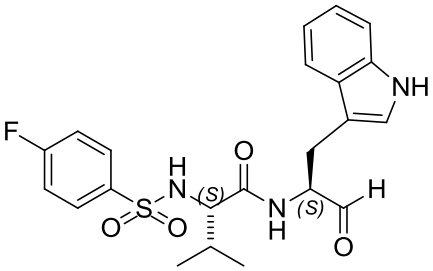
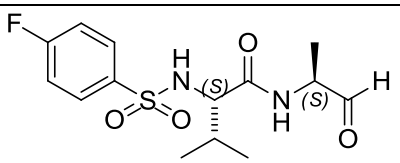
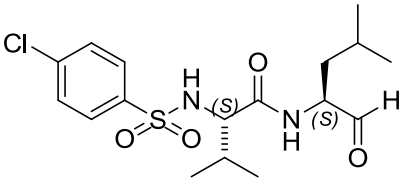
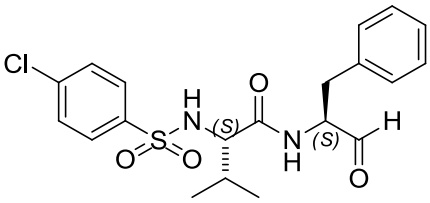
The potent calpain inhibitor SJA-6017 was first synthesized by Inoue *et al*[11] at Senju Pharmaceuticals. Their published work also reported a number of other compounds they prepared along with measured IC_{50} data against rat μ-calpain using a BODIPY fluorescent microplate assay.[11] IC_{50} is the half maximal inhibitory concentration and measures how much

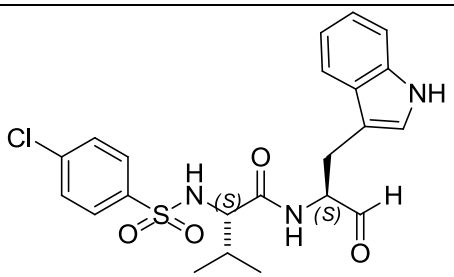
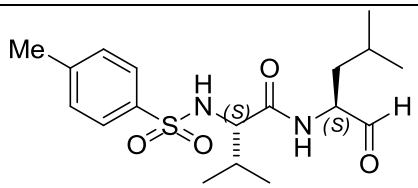
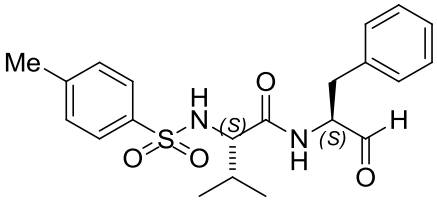
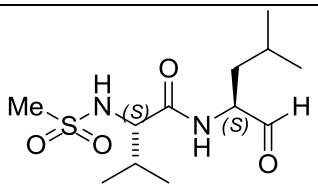
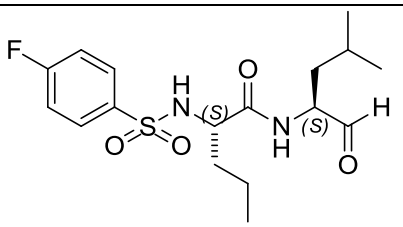
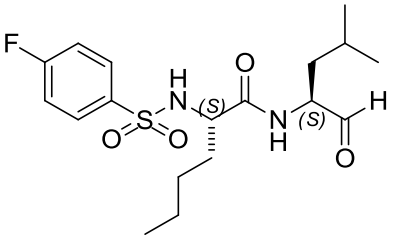
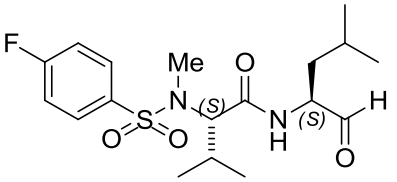
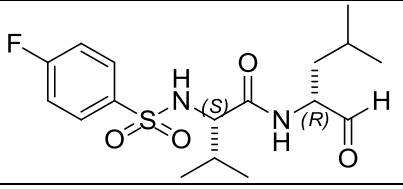
of an inhibitor is needed to inhibit a biochemical process by half. Our research group also uses the same BODIPY fluorescent microplate assay for measuring activity of cysteine proteases.[12] (see **Protocol 9** in the **Appendix** for description of BODIPY assay)

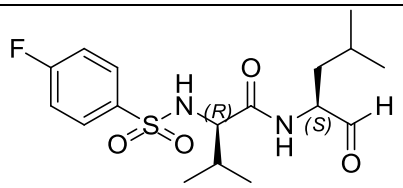
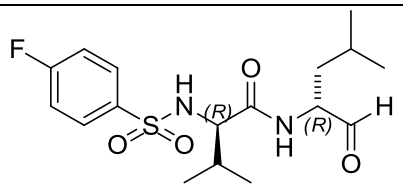
The first docking studies were performed for Inoue's eighteen compounds[11] for two reasons. Firstly, the calpain model developed for docking studies required testing to see if it was an appropriate model, one that could be used to assess potential inhibitors and thereby be used in modelling studies as a guide to the synthetic chemists in our group in choosing new compounds to synthesize. Secondly, the Inoue inhibitors were a series of compounds with a common framework but with subtle changes at various locations and with IC_{50} values that showed considerable variation with structure. Some were good inhibitors, such as SJA6017 (compound **8** in **Table 3.1**), while others were only moderate and others were very poor inhibitors of μ-calpain. This made this series of compounds an excellent starting point for modelling studies to assess whether the modelling reflected how changes in structure can affect binding to the active site of calpain.

The measured IC_{50}'s for Inoue's eighteen compounds are shown in **Table 3.1**. The main differences in structure are in the address region where different capping groups are employed, in the β-strand region where changes are made to side chains and in the configurations (L and D) of the constituent amino acids. In all cases the warhead is aldehyde.

Table 3.1: Structures and inhibitory concentrations (IC_{50}) for 18 compounds from Inoue *et al.* [11]

Compound Number	Compound Structure	IC_{50} (nM) μ-calpain
3.8		10
3.9		14
3.10		7.5
3.11		27
3.12		23
3.13		630
3.14		31
3.15		14

3.16		13
3.17		28
3.18		18
3.19		830
3.20		130
3.21		260
3.22		21000
3.23		14000

3.24		42000
3.25		1000000

3.2.2 Docking results for the Inoue *et al* compounds

The compounds **3.8**-**3.25** were first built in MacroModel and run through a Monte Carlo conformational search routine using **Protocol 3 in the Appendix**. The lowest energy conformer for each compound that displayed a *β*-strand was then used in initiation of the docking studies. These conformers were docked into the active sight of the calpain model (Chapter 2) with Glide docking software using **Protocol 4 in the Appendix**.

The docking parameters for the best pose[a] (out of a possible 10 poses generated and collected by Glide) and the inhibitory concentrations (IC_{50}) for the eighteen compounds of Inoue *et al*. [11] are shown in **Table 3.2**.

Table 3.2: Docking data for best pose[a] (out of possible 10) and inhibitory concentrations (IC_{50}) for compounds **3.8**-**3.25** from Inoue *et al*. [11]

Compound	Glide Score [b]	emodel Score [c]	Essential H bonds	Warhead Distance Å	Internal contacts			IC_{50} (nM) μ-calpain
					Good	Bad	Ugly	
3.8	-4.6	-52.1	3	3.9	200	9	0	10
3.9	-5.4	-60.6	3	3.7	197	14	0	14
3.10	-5.8	-49.4	3	4.2	203	9	0	7.5
3.11	-5.9	-54.7	3	4.0	205	14	0	27

3.12	-4.6	-59.3	3	3.6	195	7	0	23
3.13	-4.8	-49.9	3	4.1	139	10	1	630
3.14	-4.5	-45.9	3	4.0	189	11	1	31
3.15	-4.9	-55.5	3	3.7	175	10	0	14
3.16	-6.1	-62.4	3	3.4	212	10	0	13
3.17	-5.6	-50.7	3	3.9	216	18	1	28
3.18	-6.0	-54.7	3	4.9	205	9	0	18
3.19	-5.9	-48.2	3	3.7	188	9	2	830
3.20	-4.6	-49.2	2	4.0	188	20	3	130
3.21	-4.0	-52.7	2	3.6	212	15	0	260
3.22	-2.4	-45.9	1	4.4	217	15	0	21000
3.23	-4.6	-52.0	3	>5	191	15	1	14000
3.24	-3.4	-50.9	3	4.3	233	21	10	42000
3.25	-4.6	-46.3	2	3.8	206	13	5	1000000

[a] Best pose chosen by criteria in order of importance; 1 - presence of three essential hydrogen bonds, 2 - warhead within 5Å of nucleophilic Cys115, 3 - low Glide score/emodel score, 4 - lowest number of internal ugly contacts.

[b] GlideScore[13] is a scoring function based on ChemScore[14] which is designed to estimate free energy of binding for protein–ligand complex. The function uses simple contact terms to estimate lipophilic and metal–ligand binding contributions, a simple explicit form for hydrogen bonds and a term which penalises flexibility.

[c] Emodel[13] is a model energy score that combines energy grid score, binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy.

Compounds **3.8** and **3.9** are potent inhibitors of μ-calpain. The docking data for these compounds demonstrates why they should be good inhibitors; they have low negative Glide scores and Emodel scores (see description at bottom of **Table 3.2**), have the three essential hydrogen bonds for appropriate active site binding, and have a distance between the warhead and nucleophilic cysteine sulphur of less than 5Å. They also have no impossibly bad internal contacts (termed 'ugly' contacts in the software program Maestro) and is where the van der Waals contact distance is within 0.75Å. The docked poses of **3.8** and **3.9** are shown in **Figure 3.6** with the backbones (depicted as thick tubes) in the preferred β-strand conformation and oriented to form the three essential hydrogen bonds, one with Gly 271 and two with Gly 208. The naphthalene capping group of each are pointing out of the pocket and sitting over the Gly 208 and Cys 209 residues. This is similar to how the inhibitor leupeptin (**3.1**) binds to the μ-calpain construct as is seen in the PDB X-ray structure 1TL9 shown in **Figure 3.3**.[5]

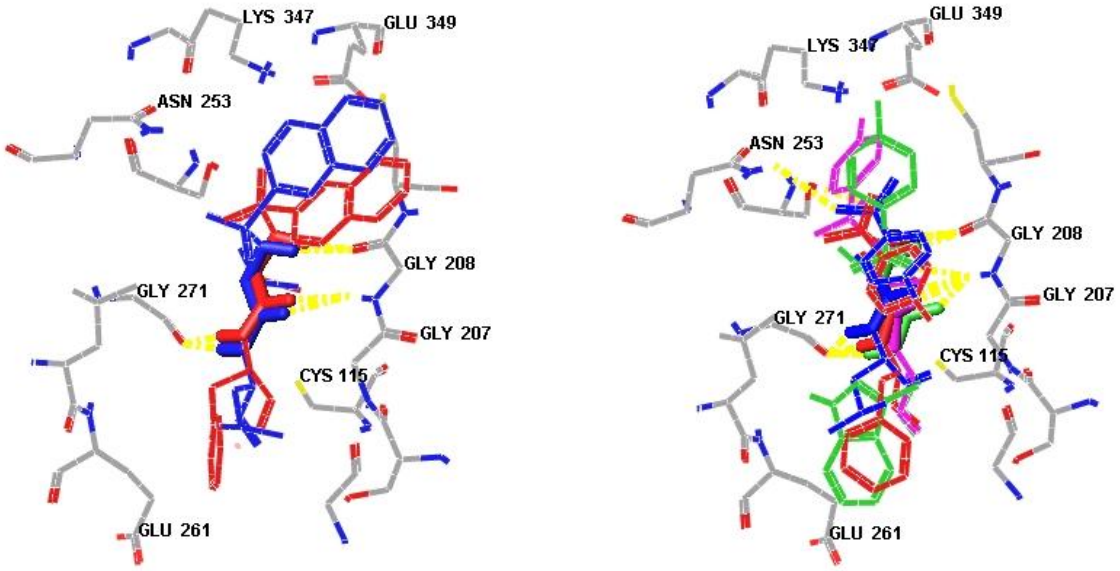


Figure 3.6: Left – best docked poses of compounds **3.8** (blue) and **3.9** (red). Right - best docked poses of compounds **3.10** (blue), **3.11** (red), **3.12** (green), and **3.13** (pink).

Compounds **3.10**, **3.11**, **3.12**, and **3.13** dock into the calpain model in a similar manner to compounds **3.8** and **3.9** (**Figure 3.6** right) but with different orientations of the 4-fluorophenyl sulfonyl capping groups. The capping groups of **3.10** and **3.11** are pointing back over the β-strand backbone of the molecule, and in **3.12** and **3.13** the capping groups are pointing towards Lys 347. Compounds **3.10**, **3.11**, and **3.12** are also potent inhibitors of calpain and the docking data suggests why this is so; they have low negative Glide Scores and Emodel Scores, have the three essential hydrogen bonds, and have a distance between the warhead and nucleophilic cysteine sulphur of less than 5Å. Note Compound **3.10** (SJA6017) was made and tested by our group against ovine m-calpain (o-CAPN2) and ovine μ-calpain (o-CAPN1). The IC_{50}'s were recorded as 80nM (m-calpain (o-CAPN2)) and 130nM (μ-calpain (o-CAPN1)) Inoue *et al* reported a value of 7.5 nM against rat μ-calpain.[11]

Compound **3.13** has a poorer IC_{50} than compounds **3.8**-**3.12**. The main difference between **3.13** and compounds **3.8**-**3.12** is the size of the side chain at the P_1 position. Compound **3.13** has an alanine at the P_1 position and the methyl side chain of this alanine has less potential for lipophilic interactions than the other compounds which have larger nonpolar sidechains at the P_1 position. In fact compound **3.13** had the highest (less favourable) energy calculated for lipophilic interactions for all compounds in the series (data not shown). The P_1 side chains are in close proximity with Leu260 and are well placed to form lipophilic interactions with the nonpolar sidechain of Leu260. The larger P_1 sidechains of all the other compounds also allow less movement of the compound within the active site. The best fit pose of compound **3.13** also has one 'ugly' internal contact.

An 'ugly' internal contact means that the molecule has had to be forced into a slightly unfavourable conformation in order to fit into the active site. Within the confines of the Glide docking program the enzyme model is kept rigid and in reality the enzyme may be able to move to accommodate such a ligand so that the ligand may in fact fit without any unfavourable internal contacts. However, such a movement in the active site of the enzyme may cause the binding energy to be higher compared with ligands that do not require such an active site enzyme movement. The reverse can also be true such that an active site movement may be required for more optimal binding of a particular ligand.

Compounds **3.14**, **3.15**, and **3.16** are good inhibitors of μ-calpain and have similar docking data as the other potent inhibitors discussed so far. They all have a 4-chlorophenyl sulfonyl capping group with their differences being at their P_1 position. Compound **3.14** has a slightly inferior IC_{50} to compounds **3.15** and **3.16** and this may result from the 'ugly' internal contact observed in the docking study.

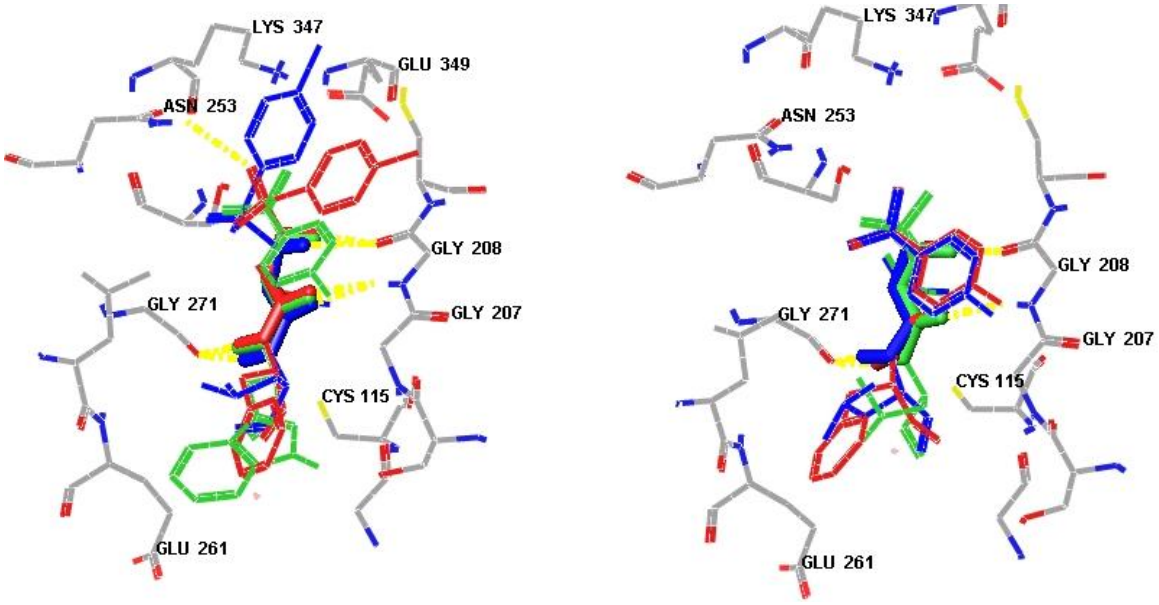


Figure 3.7: Left – best docked poses of compounds **3.14** (blue), **3.15** (red), and **3.16** (green). Right - best docked poses of compounds **3.17** (blue), **3.18** (red), and **3.19** (green).

Compounds **3.17**, **3.18**, and **3.19** can be seen docked in **Figure 3.7**. When compared with other compounds discussed above they have comparable low negative Glide scores and Emodel scores, the same three essential hydrogen bonds, and have a small warhead-nucleophile distance. Compound **3.18** has the lowest IC_{50} of these three and the docking study shows no negative data. Compound **3.17** has a slightly higher IC_{50} which may be a consequence of the 'ugly' internal contact predicted in the docking study. Compound **3.19** has an IC_{50} that is more than 100 fold higher than compound **3.10** with the data showing two 'ugly' internal contacts. It also has a small methyl sulphonamide capping group instead of a large capping group (eg 4-fluorophenyl-sulfonamide) that is seen in all the other compounds. This smaller non-aromatic capping group has less ability to form favourable interactions with the enzyme pocket than the larger capping groups which also have halogens that can form H-bonds.

Compounds **3.20** and **3.21** have L-Nval and L-Nle at P_2, respectively, as opposed to all the previously discussed compounds which have L-Val at this position. When these compounds dock into the active site analogously to known potent dipeptides of calpain,[5-7] the P_2 position fits into a deep hydrophobic pocket (S_2) in the active site cleft. L-Val seems to fit this pocket well. L-Nval and L-Nle have longer side chains than L-Val and when these compounds are docked they don't fit the pocket as well and alter the position of the *β*-strand backbone in relation to the hydrogen bond forming Gly271 and Gly208 residues. Consequently both **3.20** and **3.21** have only two of the three essential hydrogen bonds and compound **3.20** has three 'ugly' internal contacts possibly accounting for why these two compounds have IC_{50} values of only 130 and 260nM, respectively.

Compound **3.22** is structurally similar to the most potent compound **3.10** except a methyl group replaces the hygrogen at the **R** position (**Table 3.1**). When **3.10** and most of the other compounds are docked, the hydrogen at the **R** position forms one of the essential hydrogen bonds, specifically the bond to the backbone oxygen of Gly208. The substitution of the hydrogen for a methyl group removes in **3.22** the potential of the ligand to form this hydrogen bond. The methyl group is also bulkier and forces the ligand to twist in order to fit into the pocket as can be seen in **Figure 3.8**. As a result only one of the three essential hydrogen bonds is formed and the compound has a comparatively poor Glide Score of -2.4 and consistent with the modelling a poor IC_{50} of 21000 nM confirming the importance of the hydrogen bond in this case blocked by the methyl.

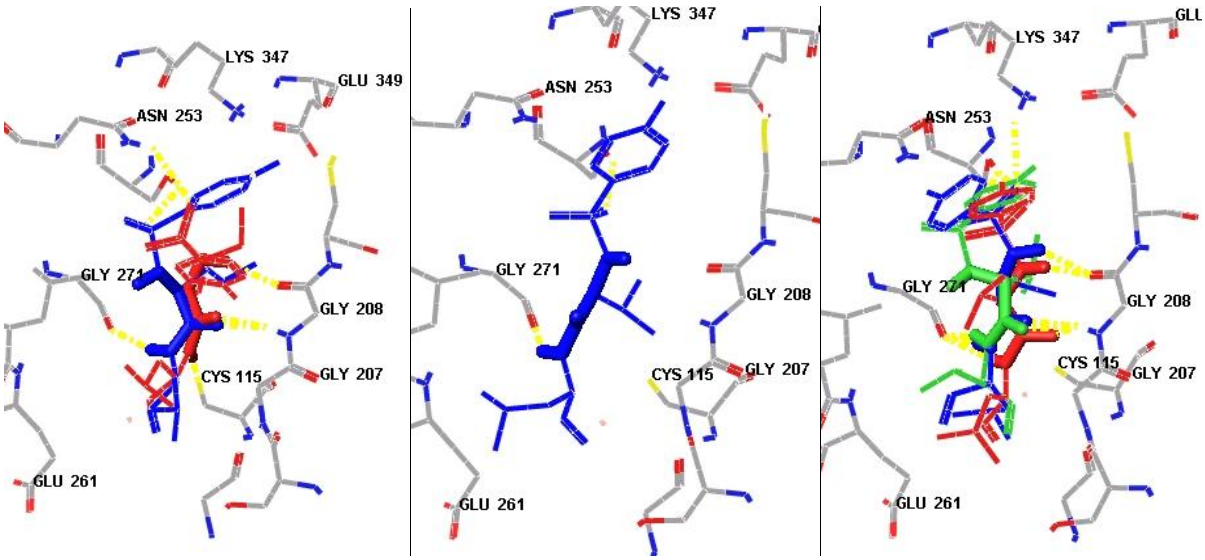


Figure 3.8: Left – best docked poses of compounds **3.20** (blue) and **3.21** (red). Middle - best docked pose of compound **3.22** (blue). Right - best docked poses of compounds **3.23** (blue), **3.24** (red), and **3.25** (green).

Compounds **3.23**, **3.24**, and **3.25** all have very poor IC_{50} values. These three compounds are the same as **3.10** except one or both amino acids have a D configuration instead of the normal L. Not surprisingly this difference in chirality has detrimental effects for ligand binding. For **3.23** a D-Leu at the P_1 position means that even with the three essential hydrogen bonds in place such a configuration results in the electrophilic aldehyde warhead not being close to the nucleophilic cysteine sulphur and accordingly being unable to be involved in a reversible covalent bond.

Compound **3.24** has valine at the P_2 postion in a D configuration. The docking studies show that this compound could dock with the three appropriate hydrogen bonds and with a close warhead-cysteine distance but not without having ten 'ugly' internal contacts. Most of these 'ugly' contacts are due to the P_1 and P_2 side chains being impossibly close to each other, and the P_2 side chain in an impossible proximity to the phenyl ring. The docked pose of **3.24** in **Figure 3.8** also shows that the side chain of the D-Val is pointing out of the active site

pocket instead of deep in the hydrophobic region of the cleft where the P_2 side chain is found in poses of the more potent dipeptide ligands. The loss of this seemingly crucial hydrophobic contact is reflected by the poorer Glide score of -3.4.

Compound **3.25** has both the P_1 and P_2 amino acids in a D configuration. The best docked pose of this compound shows that it can only form two of the three required hydrogen bonds when the warhead is in a position for nucleophilic attack. In order to achieve this pose five 'ugly' internal contacts form, most of which involve the close proximity of the P_2 sidechain and the capping group end which is an unfavourable conformation.

Of the eighteen compounds docked, eight compounds (**3.8**, **3.9**, **3.10**, **3.11**, **3.12**, **3.15**, **3.16**, and **3.18**) have an IC_{50} of less than 30nM against μ-calpain. When docked these eight have poses that show a warhead-cysteine sulphur distance of less than 5Å, which is an appropriate distance necessary for nucleophilic attack to occur on the aldehyde and thus be in a position for a reversible covalent bond to form. They all meet the requirement of having the three essential hydrogen bonds, an important constraint of a β-strand moiety, a conformation observed in almost all known protease-ligand complexes.[13] Their P_1 side chains form hydrophobic contacts with Leu260 and their P_2 side chains fit deep into the hydrophobic pocket of the active site. All display poses that contain no 'ugly' internal contacts, indicative of not being forced into an impossible conformation in order to fit into the pocket. The capping groups of these eight compounds are found to be all pointing out of the pocket in variable positions in relation to the rest of their structure; some are folded back over the dipeptide backbone, some are over Gly 208, and others are pointing away from the dipeptide backbone.

The variable positioning of the capping groups is consistant with the findings of Qian *et al* where the electron density around the capping group of two α-ketoamide inhibitors was found to be weak within their solved X-ray crystal structures (2R9C and 2R9F), suggesting that they are flexible and form limited interactions with the S_3 unprimed region.[7] The X-ray structures of these two α-ketoamides show their capping groups sitting over Gly208 as was observed in the docking of several of the Inoue *et al*[11] compounds. The protecting group of the two α-ketoamides is carboxybenzyl which has no substituents on the aromatic ring that can act as hydrogen bond donors or acceptors. In contrast all the Inoue *et al*[11] compounds with the exception of compounds **3.8**, **3.9**, and **3.19** have either a fluorine or chlorine substituent on their capping group, with the the potential to form hydrogen bonds. These hydrogen bonds are possible at the P_3 subsite residues and may possibly form between the capping group halogens and residues Lys347 and Asn253 with a small flexible movement of these active site residues. Their GlideScores are between -4.6 and -6.0.

The other ten compounds have at least one of the above mentioned criteria that are less than satisfactory for binding (highlighted yellow in **Table 3.2**) and this modelling correlates with their inferior IC_{50} results.

3.3 Docking studies of SJA-6017 analogues

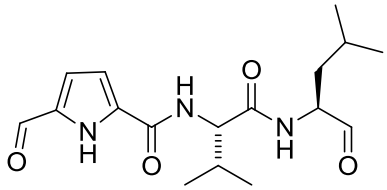
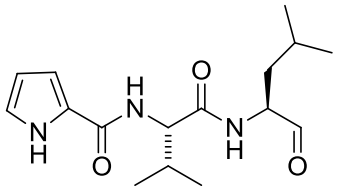
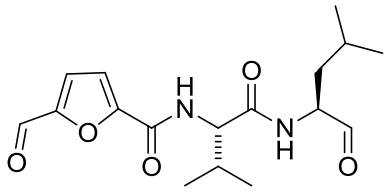
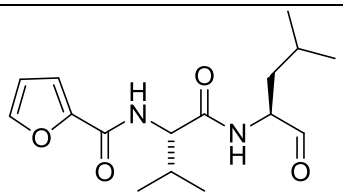
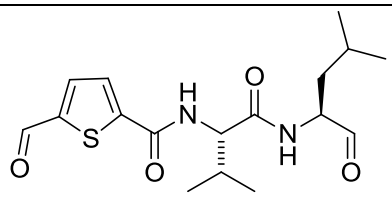
3.3.1 *N*-Heterocyclic dipeptides

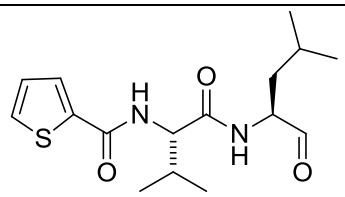
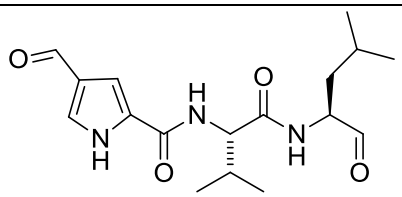
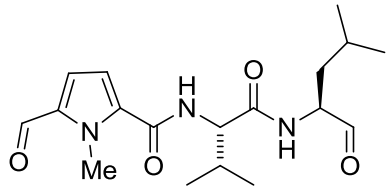
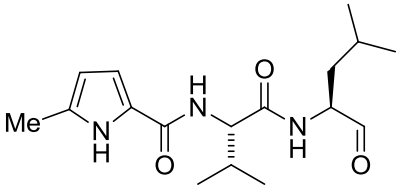
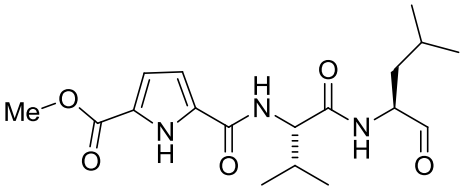
Our calpain research group wanted to investigate different capping groups at the P_3 position to see if additional binding could be achieved in the S_3 region of calpain. A series of

N-heterocyclic analogues of SJA-6017 were suggested by the synthetic chemists in our research group and molecular modeling studies were used to first assess their potential as calpain inhibitors and to rationalize any differences in IC_{50} values measured for these compounds where it was decided it was worth synthesizing.

Ten *N*-heterocyclic dipeptides that were synthesized along with their corresponding measured IC_{50} results are shown in **Table 3.3**. The dipeptides have an aldehyde warhead and the same backbone as the lead compound SJA6017. The *N*-heterocyclic capping group varies at the X, R_1, and R_2 positions.

Table 3.3: Structures and inhibitory concentrations (IC_{50}) for ten *N*-heterocyclic dipeptides (**3.26**-**3.35**)

Compound Number	**Compound Structure**	**IC_{50} (nM) o-CAPN1**	**IC_{50} (nM) o-CAPN2**
3.26		290	25
3.27		650	315
3.28		960	100
3.29		790	135
3.30		440	85

3.31		680	100
3.32		530	100
3.33		150	150
3.34		340	110
3.35		290	140

3.3.2 Docking results of *N*-heterocyclic dipeptides

The ten dipeptides were built in MacroModel[14] and a conformational search of each was undertaken using **Protocol 3 in the Appendix**. The lowest energy conformer for each compound that displayed a *β*-strand was used as a starting structure for the docking study and were identified by Xcluster analysis.[15] These low energy conformers were docked into the active sight of the calpain model with Glide docking software using **Protocol 4 in the Appendix**. The low energy conformers for the ten dipeptides were also used to dock into the calpain model using the InducedFit docking program using **Protocol 5 in the Appendix**. The

InducedFit Protocol was performed to compare the results with the Glide docking results to determine if InducedFit was a better predictive tool than Glide docking.

The Glide docking data for the best pose (out of a possible 10 poses generated and kept by Glide) and the inhibitory concentrations (IC_{50}) against ovine calpain 1 (o-CAPN1) and ovine calpain 2 (o-CAPN2) for the ten compounds are shown in **Table 3.4**.

Table 3.4: Glide docking data for best pose (out of a possible 10) and inhibitory concentrations (IC_{50}) of ten *N*-heterocyclic dipeptides.

	Glide	emodel	Essential	Warhead	Internal contacts			IC_{50} (nM)	
Cpd[a]	score	score	H bonds	distance	Good	Bad	Ugly	o-CAPN1	o-CAPN2
3.26	-6.2	-54.7	3[b]	3.6	197	20	3	290	25
3.27	-5.0	-49.3	3.0	4.0	177	4	2	650	315
3.28	-5.0	-52.9	3.0	4.3	181	5	1	960	100
3.29	-5.6	-48.4	3.0	3.7	173	5	0	790	135
3.30	-5.3	-53.8	3.0	4.1	189	5	0	440	85
3.31	-5.0	-46.6	3.0	3.9	177	3	0	680	100
3.32	-5.1	-49.9	3.0	4.5	184	8	2	530	100
3.33	-5.3	-49.9	2[b]	3.7	199	17	2	150	150
3.34	-5.4	-46.7	0.0	3.5	180	9	0	340	110
3.35	-3.2	-49.8	3.0	3.7	187	6	0	290	140

[a] Compound.
[b] Hydrogen bonds from the Gly271 carbonyl and Gly208 carbonyl and NH group to the ligand are present but not bonded to the usual ligand donor and acceptor groups of the ligand backbone.

Compound **3.26** is potent against m-calpain (o-CAPN2) and is >11 fold more selective for m-calpain over μ-calpain (o-CAPN1). The docked pose of **3.26** on the left hand side of **Figure 3.9** has the best Glide score of -6.2 and has a different binding pattern compared with other potent dipeptide inhibitors and this may explain why it is more potent than other inhibitors in this series. The essential hydrogen bond formed between Gly271 and the NH at the P_1 position is the same as that seen in other potent dipeptides, however, hydrogen bonds between Gly208 and the NH and carbonyl group at the P_2 position are not observed. Instead the ligand, in comparison with other potent dipeptides, is twisted so that hydrogen bonds are formed between the carbonyl at the P_3 peptide bond and the backbone NH of Gly208, and

between the NH of the pyrrole and the backbone carbonyl of Gly208. The warhead is still held in a good position and orientation for nucleophilic attack by the sulphur of Cys115. There is an additional hydrogen bond between the P_1 aldehyde and the side chain of Gln109.

In spite of this unique hydrogen bonding formation, the compound loses some hydrophobic contacts because of the rotation necessary for the hydrogen bonds to form while still having a close warhead-nucleophile distance. This causes the valine sidechain to point out of the cleft as opposed to positioning itself deep into the hydrophobic pocket. It also produces three 'ugly' contacts since the sidechains of valine and leucine are brought close together.

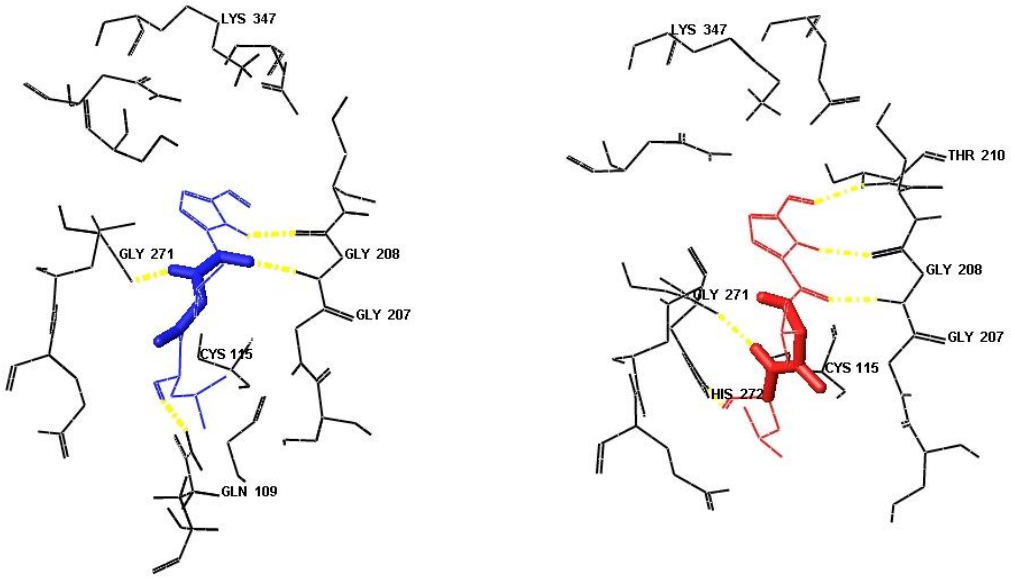


Figure 3.9: Left – best docked pose of compound **3.26** (blue) from Glide docking. Right - best docked pose of compound **3.26** (red) from the InducedFit docking. Hydrogen bonds shown as dashed yellow lines.

When compound **3.26** was docked using the InducedFit protocol (right hand pose in **Figure 3.9**) the result was a similar pose. The pose shows the same general orientation as that seen on the left of **Figure 3.9** (Glide docking pose). In addition a hydrogen bond is formed between the aldehyde at the pyrrole end of the compound and the NH backbone of Thr210.

The bond between the P_1 aldehyde and the side chain of Gln109 is lost and replaced with one between the P_1 aldehyde and the side chain of His272 due to a slight change in position of the aldehyde within the pocket but does not change its close proximity to the nucleophile. The InducedFit pose still has three 'ugly' internal contacts. InducedFit seemingly has no benefit over Glide docking.

Compound **3.27** is one of the least potent in the series and is only twice as selective for m-calpain (o-CAPN2) over μ-calpain (o-CAPN1). It forms the three essential hydrogen bonds when docked into the μ-calpain model using Glide and forms two additional hydrogen bonds when docked with the InducedFit Protocol (**Figure 3.10**). One additional bond is formed between the carbonyl at the capping group end and the side chain of Asn253. The other is between the NH of the pyrrole and the backbone of Gly208. It is not clear from the docking studies why this is not as potent as others in the series except for two 'ugly' internal contacts.

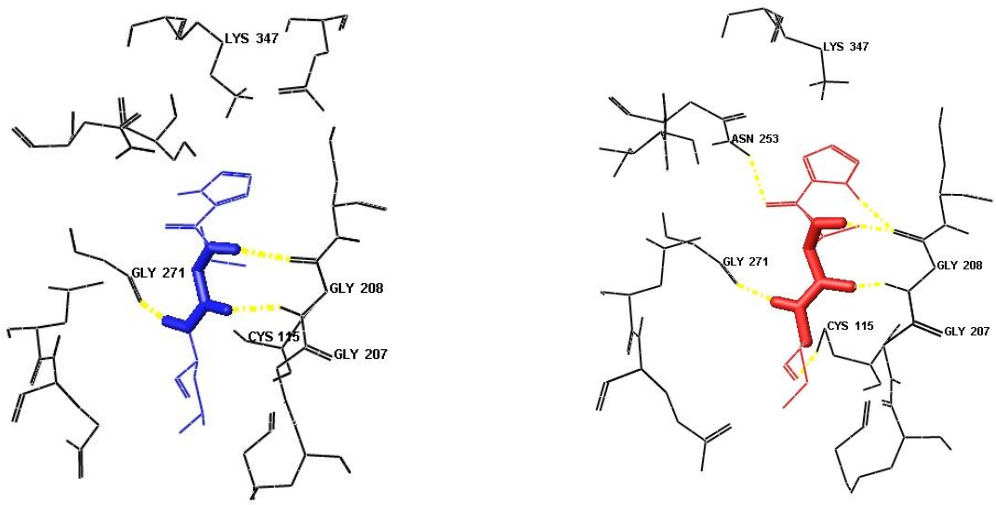


Figure 3.10: Left – best docked pose of compound **3.27** (blue) from Glide docking. Right - best docked pose of compound **3.27** (red) from the InducedFit docking. Hydrogen bonds shown as dashed yellow lines.

The left picture in **Figure 3.11** shows the best docked poses of compounds **3.28**, **3.29**, **3.30**, and **3.31**. These four all dock in a similar manner with all the essential hydrogen bonds in place with a close warhead-nucleophile proximity. Their IC_{50}'s are similarly potent against m-calpain (o-CAPN2) at around 100nM, but vary against μ-calpain (o-CAPN1). Compound **3.28** (blue) and compound **3.31** (pink) have only the three essential hydrogen bonds in their respective best poses. Additional hydrogen bonding is found in the docked poses of compounds **3.29** and **3.30** from their respective capping groups to the enzyme. **3.29** (red) displays a hydrogen bond from the furan ring oxygen to the side chain of Asn253. The sulphur, in the thiophene ring of compound **3.30** (green), forms a hydrogen bond to Lys347 and another from the aldehyde to the side chain of Lys347. These additional hydrogen bonds are favourable and would contribute to tight binding of the compounds to the active site. Compounds **3.28** and **3.31** also display hydrogen bonding from their capping group oxygen and sulphur, respectively, in other docked poses (not shown).

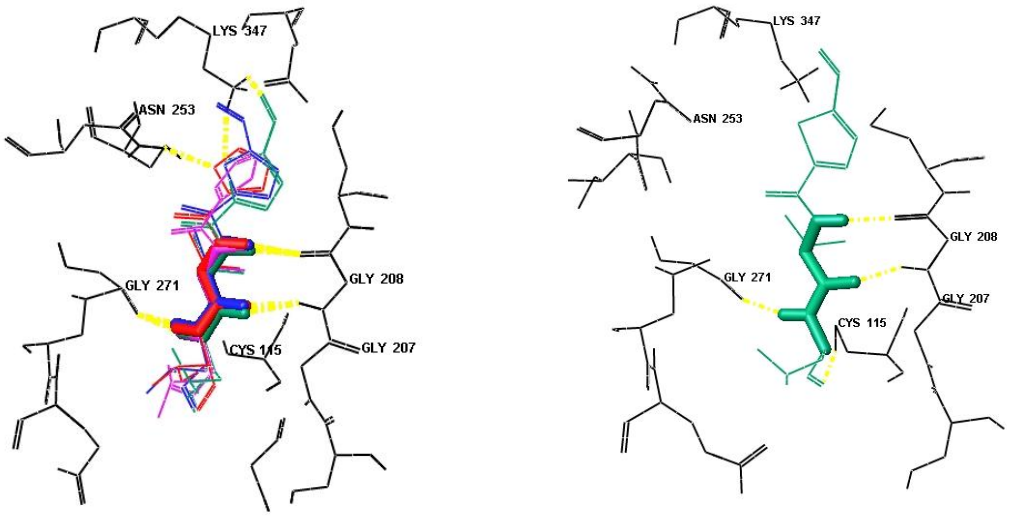


Figure 3.11: Left – best docked pose of compound **3.28** (blue), **3.29** (red), **3.30**, (green) and **3.31** (pink) from Glide docking. Right - best docked pose of compound **3.30** (green) from the InducedFit docking, Hydrogen bonds shown as dashed yellow lines.

When the InducedFit protocol was used compounds **3.28**, **2.29**, and **3.31** did not return any docked poses. Compound **3.30** returned one docked pose shown on the right of **Figure 3.11**. However, the two additional hydrogen bonds seen in the Glide docked pose are not observed in the InducedFit pose as the side chain of Lys347 has moved. InducedFit modelling for these compounds has not been beneficial.

Compound **3.32** shown below in **Figure 3.12** (left) docked (with Glide) into the active cleft and shows the three essential hydrogen bonds in place and the warhead in position for nucleophilic attack. It is one of the most potent of compounds in the series against m-calpain (o-CAPN2) but is five times less potent against μ-calpain (o-CAPN1). There is one extra bond involving the pyrrole aldehyde and the side chain of Lys347. When docked with InducedFit this extra hydrogen bond is lost due to movement of the lysine residue.

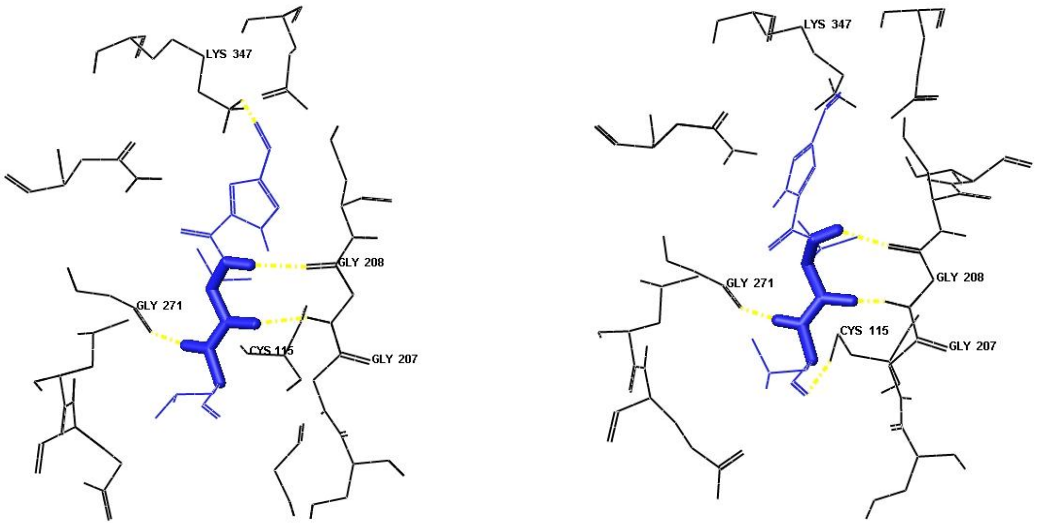


Figure 3.12: Left – best docked pose of compound **3.32** (blue) from Glide docking. Right - best docked pose of compound **7** (blue) from the InducedFit docking. Hydrogen bonds shown as dashed yellow lines.

Compound **3.33** docked into the active site with Glide (**Figure 3.13**) in an analogous manner to that of compound **3.26**. Compound **3.33** and **3.26** have similar structures - a methyl group in compound **3.33** replaces the NH hydrogen in the pyrrole ring of compound **3.26**. The methyl group in **3.33** does not allow for a hydrogen bond to form, as occurs for **3.26** between the NH of the pyrrole and the backbone oxygen of Gly208. The methyl group being bulkier than the hydrogen also causes the pyrrole ring to twist in order to fit into the active site pocket. A H-bond does form between the pyrrole aldehyde and Lys347. This compound is the most potent against i-calpain (o-CAPN1) at 150nM but is unselective as it also has an IC_{50} of 150nM against (o-CAPN2). No pose was returned when compound **3.33** was run through the InducedFit protocol.

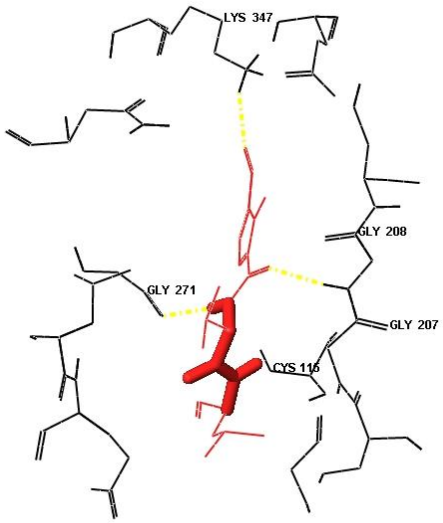


Figure 3.13: Left – best docked pose of compound **3.33** (red) from Glide docking. Hydrogen bonds shown as dashed yellow lines.

The pose on the left of **Figure 3.14** shows compound **3.34** docked into the active site using Glide. As depicted the compound does not form any of the essential hydrogen bonds seen in other potent dipeptide inhibitors.

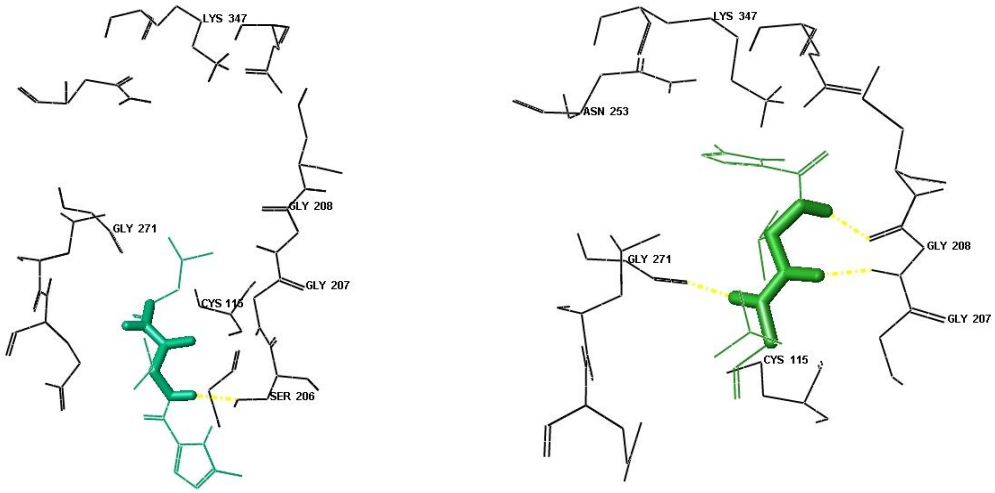


Figure 3.14: Left – best docked pose of compound **3.34** (green) from Glide docking. Right - best docked pose of compound **3.34** (green) from the InducedFit docking. Hydrogen bonds shown as dashed yellow lines.

On the right of **Figure 3.14** is the best docked pose of compound **3.34** using the InducedFit Protocol. This pose is more typical of a potent dipeptide calpain inhibitor with it having the three essential hydrogen bonds. The capping group is unusually positioned such that it is twisted and pointing towards the Asn253 residue. The capping group does not form any additional hydrogen bonds with the enzyme. The *in vitro* testing revealed **3.34** to be reasonably potent against m-calpain (o-CAPN2) with an IC_{50} of 110nM, three times more potent than against μ-calpain (o-CAPN1). InducedFit docking of compound **3.34** was nessessary as Glide docking did not produced any viable poses.

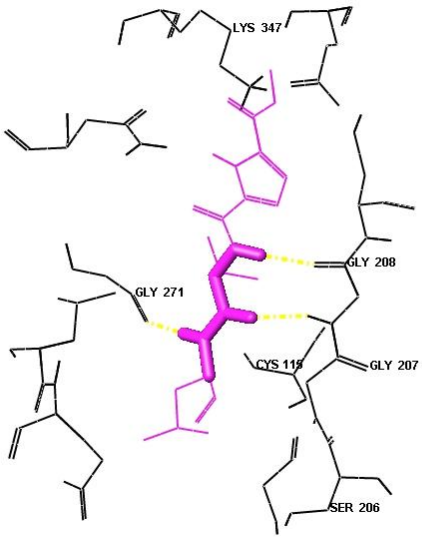


Figure 3.15: Left – best docked pose of compound **3.35** (pink) from Glide docking. Hydrogen bonds shown as dashed yellow lines.

Compound **3.35** displays typical hydrogen bonding for a potent dipeptide and the aldehyde warhead is in position for nucleophilic attack (**Figure 3.15**). It inhibits m-calpain (o-CAPN2) with an IC_{50} of 140nM and μ-calpain (o-CAPN1) with an IC_{50} of 290nM. The capping group displays no hydrogen bonding, although it is in a position where such bonding could occur if small movements in the surrounding residues facilitated this happening. There was no InducedFit poses returned for this structure.

In summary, the ten *N*-heterocyclic dipeptides (compounds **3.26**-**3.35)** are all potent or semi-potent calpain inhibitors as shown by their low to reasonably low IC_{50}'s against both m-calpain (o-CAPN2) and μ-calpain (o-CAPN1). Compounds **3.26**, **3.28**, and **3.31** are highly selective for m-calpain (o-CAPN2) over μ-calpain (o-CAPN1) while the rest display only moderate to little selectivity.

All the compounds, excluding **3.26**, **3.33** and **3.34**, dock using Glide in an extended β-strand conformation with the three essential hydrogen bonds intact and with adjacent

warhead-nucleophile proximity. Compounds **3.28**-**3.32** all have hydrogen bond acceptors in their capping groups that form bonds with donators from the side chains of either Asn253 or Lys347. These hydrogen bonds are seen in either the best poses of each docked structure or similar poses within the top ten poses that Glide produced. Compound **3.27** has no hydrogen bond acceptor in its capping group and this may reflect why it is the least potent against m-calpain (o-CAPN2) and one of the least potent against μ-calpain (o-CAPN1).

Compound **3.34** docks in an appropriate manner only using the InducedFit docking. Like compound **3.27** it has no hydrogen bond acceptors in its capping group. The 5-methyl pyrrole capping group is uniquely twisted to one side and the methyl group is pointing into a hydrophobic pocket formed by the enzyme residues Ile257 and Leu260 giving rise to a hydrophobic-hydrophobic interaction (**Figure 3.16**). This interaction could explain why **3.34** is more potent than compound **3.27** which does not have the hydrophobic methyl group.

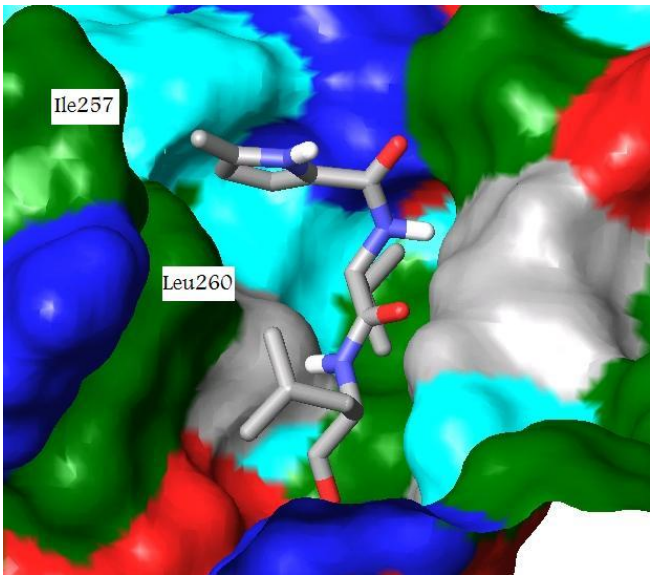


Figure 3.16: Best docked pose of compound **3.34** (tube structure) from the InducedFit docking, The enzyme is shown as a surface representation with residues colour coded by residue property; green = hydrophobic, blue = positive, red = negative, cyan = polar uncharged, and grey = glycine.

The methoxycarbonyl pyrrole group of compound **3.35** has the potential to form hydrogen bonds with the likes of Lys347 and Asn253 but no such interactions were observed in any of the poses generated by Glide. The other poses generated by Glide had the compound docking in a similar mode to that of compound **3.26**.

All ten compounds showed varying degrees of potency and selectivity although the difference between the best and worst compound in each category was not large. Most of the compounds best docked poses showed the ability or potential to form hydrogen bonds with the residues surrounding the S_3 subsite of the enzyme.

Compound **3.26** has a distinctive binding mode (compound **3.35** also showed this mode in poses not shown). In reality it may be able to bind in the more typical manner seen in the other docked compounds and has the potential to form hydrogen bonds between the pyrrole aldehyde and surrounding residues such as Lys347. The only way to determine the actual mode of binding would be to undertake an X-ray crystal structure of compound **3.26** co-crystallised with the enzyme. However, it is likely that compound **3.26** actually binds in the typical β-strand mode and that the manner in which it docks with Glide is a 'figment' of the program. It may have two distinct ways of binding the enzyme tightly, the one shown in the docking and the typical β-strand mode which could explain its higher potency. If this unusual binding mode is in fact the actual manner in which it binds then that may be the reason for greater potency alone. However, this unusual binding pattern has some 'ugly' internal contacts and the valine at the P_2 position is pointing out of the pocket instead of it pointing into the hydrophobic S_2 subsite which is more typical of potent di-peptides.

Using the InducedFit docking did not add anything to the results of the Glide docking except in the case of compound **3.34** which did not dock appropriately with Glide but did produce a viable pose with InducedFit. Therefore, it may only be nessessary to use the InducedFit Protocol on ligands of this type (di-peptide aldehydes) when Glide does not produce the expected docked poses.

3.4 Docking studies of diazo and triazene compounds

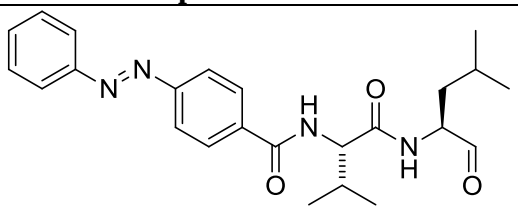
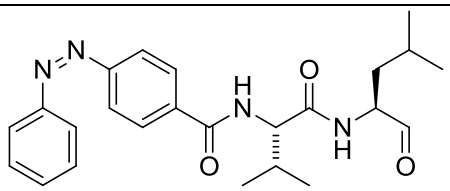
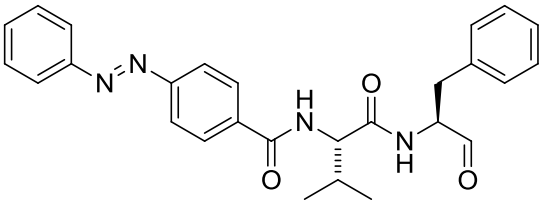
The following is work has been published in a paper on the modelling of diazo- and triazene dipeptide aldehydes;

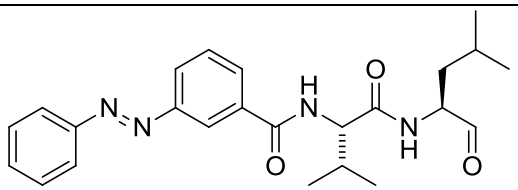
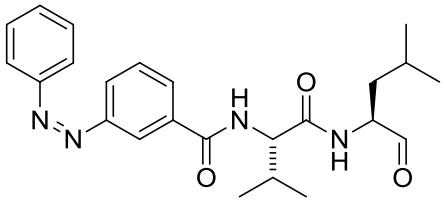
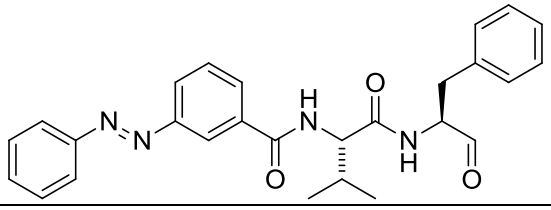
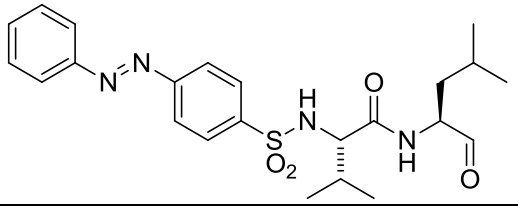
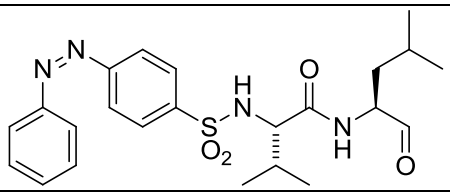
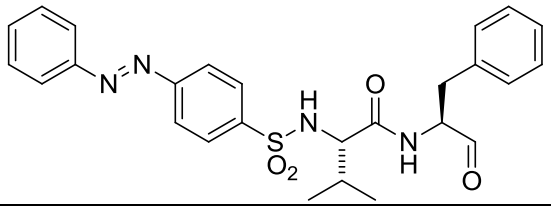
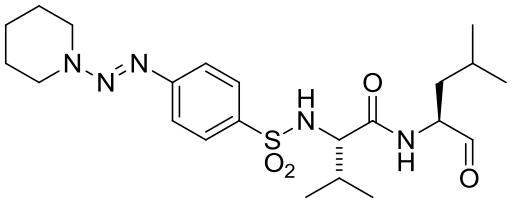
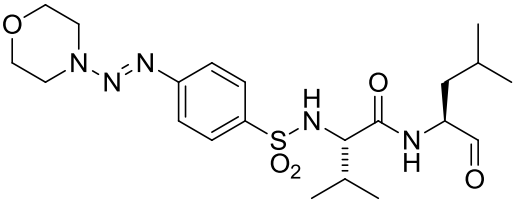
Investigation into the P3 Binding Domain of m-Calpain Using Photoswitchable Diazo- and Triazene-dipeptide Aldehydes: New Anticataract Agents
Abell, A. D.; Jones, M. A.; Neffe, A. T.; Steven G. Aitken; Cain, T. P.; Payne, R. J.; McNabb, S. B.; James M. Coxon; Stuart, B. G.; Pearson, D.; Lee, H. Y.-Y.; Morton, J. D., *J. Med. Chem.* **2007,** *50*, 2916-2920.

A series of (*E*)- diazodipeptide aldehydes (amides **3.36a-d** and sulphonamides **3.37a**,**b**) that contain an *N*-terminal diazo group designed to extend deep into the S_3 binding pocket and a C-terminal aldehyde for attachment to the active site cysteine were prepared by members of the cataract group. The *N*-terminal diazo groups of (*E*)-**3.36a**,**c** and (*E*)-**3.37a** were photochemically isomerized to give an alternative photostationary state in which the (*Z*)-isomer predominates. These mixtures, enriched in either the (*E*)-or (*Z*)-isomer, were assayed against m-calpain to further explore and define the P_3 binding pocket. This new class of *N*-terminal group was extended to more water soluble triazene-dipeptide aldehydes **3.38a**,**b**, with **3.38a** being assayed in lens culture to determine its ability to arrest the development of calpain-induced cataract.

Samples of the dipeptidic aldehydes **3.36a-d**, **3.37a**,**b**, and **3.38a**,**b**, consisting of predominantly the (*E*)-isomer, were assayed against m-calpain using a fluorescence-based assay[12] to determine *in vitro* potency, and the results are summarized in **Table 3.5**. The initial (*E*)/(*Z*)-isomer mixtures for **3.36a**, **3.36c**, and **3.37a** were irradiated with ultraviolet light (500 W mercury arc lamp through a UV filter with a narrow wavelength band centered at 340 nm) to give samples enriched in the (*Z*)-isomer [1:4.3, 1:4, and 1:3.7 mixture of (*E*)- and (*Z*)-isomers, respectively, **Table 3.5**]. These new photostationary states of **3.36a**, **3.36c**, and **3.37a** (**Table 3.5**) were also assayed against m-calpain to further assess the available space in the S_3 binding pocket, to which the diazobenzene groups purportedly bind as revealed by molecular modeling.[16] The computational methods can be found under **Protocol 6** in the **Appendix**.

Table 3.5: Inhibition of m-Calpain by Different Photostationary States of the Diazo and Triazene-dipeptide Aldehydes **3.36a-d**, **3.37a**,**b**, and **3.38a**,**b**

			IC_{50}
Compound[a]	*E*/*Z*	**Compound Structure**	**(nM) m-calpain**
3.36a	4.8:1		45
3.36a (irradiated)	1:4.3		175
3.36b	5.4:1		35

3.36c	4.9:1		75
3.36c (irradiated)	1:4		105
3.36d	5.4:1		170
3.37a	3.3:1		40
3.37a (irradiated)	1.3:7		100
3.37b	7:1		90
3.38a	23:1		90
3.38b	233:1		420

[a] The major isomer before irradiation is assigned as the thermodynamically more stable (*E*)-isomer, based on literature precedence.12 [b] Substitution position.

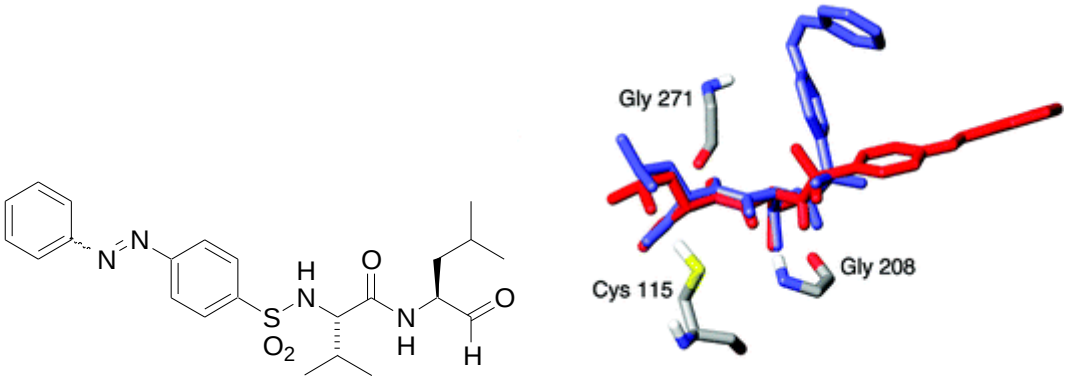


Figure 3.17: Picture of docked compound **3.37a** in the (*E*)-enriched photostationary state overlayed with **3.37a** in the (*Z*)-enriched photostationary state.

The most active dipeptidic aldehyde mixture **3.36b** had an IC_{50} value of 35 nM, which is significantly more potent than the calpain inhibitor SJA6017 **3.10** (IC50) 80 nM in our assay).[17] The inhibitor mixtures **3.36a** (45 nM), **3.36c** (75 nM), **3.37a** (40 nM), and **3.38a** (40 nM) are also particularly potent. Of all the sample mixtures, only **3.36d** (170 nM) and **3.38b** (420 nM) are less potent than **3.10** (**Table 3.5**).

The S_1 pocket can accommodate both Leu and Phe at the P_1 position; however, there does seem to be some preference for Leu at this position; compare assay results for **3.36a**,**b**, **3.36c**,**d**, and **3.37a**,**b** (**Table 3.5**). With regard to potency, there does not seem to be any consistent preference for an amide or sulfonamide linker; compare the activity of **8a**-**10a** and **3.36b**-**3.37b** (**Table 3.5**). Inhibitors with 4-substitution of the *N*-terminal aryl group (**3.36a**,**b**) are more potent than those with 3-substitution (**3.36c**,**d**).

The triazenes **3.38a**,**b** are less potent than the diazo compounds **3.36a-d** and **3.37a**,**b**. The planar and extended aromatic nature of the (*E*)-diazo group of **3.36a**,**b** and **3.37a**,**b** appears to allow for better interaction with this flat S_3 binding pocket, when compared to the

corresponding triazenes **3.38a**,**b**. This is consistent with the lower potency of the triazenes **3.38a**,**b** (see **Table 3.5**).

Induced fit modeling[18] shows each μ-calpain docked inhibitor adopts a β-strand conformation.[19] The poses of (*E*)-**3.36a**, (*E*)-**3.36b**, (*E*)-**3.36c**, (*Z*)-**3.36c,** (*E*)-**3.37a**, (*E*)-**3.38a**, and (*E*)-**3.38b** docked with a μ-calpain construct are shown in **Figure 3.18a**. This is consistent with published crystal structures of μ-calpain with inhibitor bound.[5, 6, 20] The docked poses for (*Z*)-**3.36a**, (*E*)-**3.36d**, (*Z*)-**3.37a**, and (*E*)-**3.37b** (not shown in **Figure 3.18a**) also revealed a β-strand conformation of the inhibitor but with only two of the three hydrogen bonds apparent. The model also predicts the aldehyde group in all cases to be in close proximity to the active site cysteine, as required for mechanism-based inhibition.

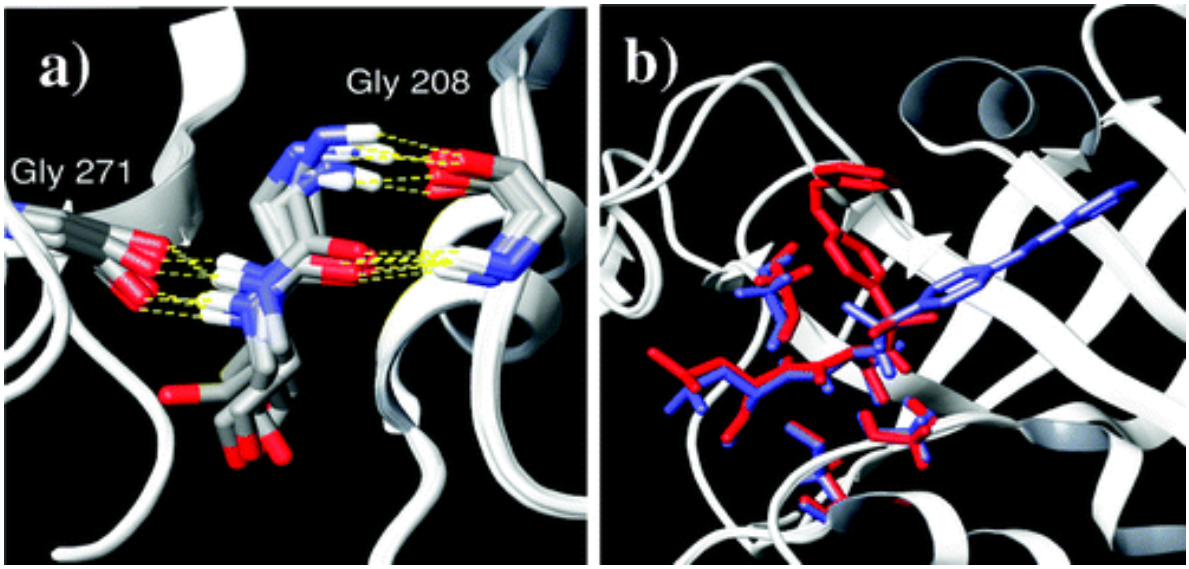


Figure 3.18: (a) β-Strand backbone conformations of (*E*)-**3.36a**, (*E*)-**3.36b**, (*E*)-**3.36c**, (*Z*)-**3.36c**, (*E*)-**3.37a**, (*E*)-**3.38a**, and (*E*)-**3.38b**, resulting from the inducedfit modeling. (b) Induced-fit overlay of the S3 subsite of calpain with docked (*E*)-**3.37a** (blue) and (*Z*)-**3.37a** (red).

The inhibitory activity of the isomeric mixtures of **3.36a**, **3.36c**, and **3.37a** decreased, on irradiation, by a factor of 3.9, 1.4, and 2.5, respectively. For example, the inhibitory activity

of **3.36a** decreased from 45 nM for a sample containing 83% (*E*)-isomer to 175 nM for a mixture containing 4-fold less of the (*E*)-isomer (i.e., 19% (*E*)-isomer with 81% (*Z*)-isomer), see **Table 3.5**. The (*E*)-isomers are significantly more active in each case,[21] reflecting a better fit in the S_3 binding pocket associated with their geometry and dipole moment. In fact, modeling suggests that the diazo group of (*Z*)-**3.37a**, unlike (*E*)-**3.37a**, appears not to bind in the S3 binding pocket, but rather interacts with a hydrophobic patch on the mobile loop that defines the calpain active site (**Figure 3.18b**). This is consistent with the observed higher potency of the photostationary state enriched in the (*E*)-isomer (see **Table 3.5** and previous discussion). A photoisomerizable *N*-terminal diazobenzene of this type provides both a useful probe for mapping the size and geometry of the S_3 binding domain and evidence for enhancing interactions between this pocket and a P_3 aryl group of an inhibitor. An empirical preference for the interaction of an aryl group with S_3 has been previously suggested,[5] and our study supports this observation.

The *in vivo* potential of these highly potent inhibitors was next determined by assessing their ability to retard calpain induced cell damage in sheep lenses cultured in Eagle's minimal essential medium. Inhibitor **3.38a** was chosen for study because of its superior water solubility, as compared to **3.36a-d** and **3.37a**,**b**, and also its high *in vitro* potency, as compared to the other triazene **3.38b**. The triazene moiety of **3.38a** is also amenable to salt formation that makes it attractive as a potential drug candidate.

Inhibitor **3.38a** (0.8 μM) was added to one lens from each pair of sheep lenses in culture media by our Lincoln collaborators. Two hours later calcium was added to all lenses to activate the constituent calpains and hence induce cataract formation. After 24 h, all lenses were photographed and the opacity graded; see **Figure 3.19** for representative examples. Lenses treated with calcium only (e.g., lens 1 in **Figure 3.19**) clearly showed the opacity

associated with cataract formation; however, lenses treated with calcium in the presence of **3.38a** (e.g., lens 2) remained essentially clear, as revealed by the reference grid placed behind each lens. The loss of transparency was significantly reduced by **3.38a** ($p < 0.005$) in a paired t-test. It would thus appear that this class of inhibitor, with an *N*-terminal group capable of binding deep into the S_3 binding pocket, is active *in vitro* and *in vivo* (lens assay) and is thus of interest as a potential drug candidate.

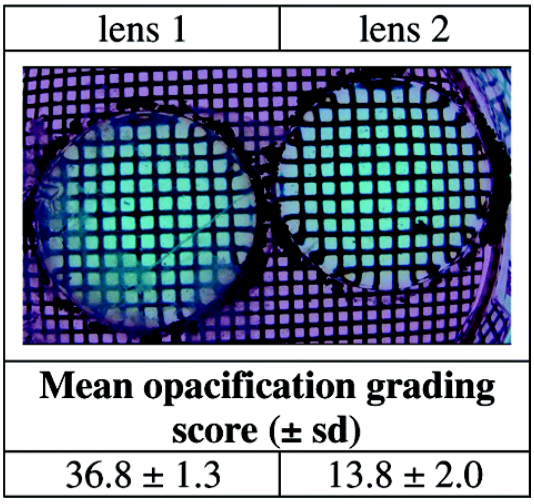


Figure 3.19: Calcium-induced cataract in sheep lenses. The scores represent the average result using three lens pairs. Opacification scores of 100 = full opacity, whereas a score of 1 = clear and transparent.

3.5 Conclusion

Known inhibitors of calpain were examined in detail to determine what attributes they possessed that determine them to be good inhibitors. Most of the inhibitors in the literature are peptide based and are known to bind in a *β*-strand conformation. The known natural inhibitors such as leupeptin were found by researchers to be poor drug candidates as they were non-selective and had low membrane permeability. Other researchers discovered that this could be overcome by the addition of *N*-terminal capping groups, a fact that lead our team to research further.

The Glide model was tested to see if it was a viable model for docking studies. The compounds of Inoue *et al*[11] were deemed to be ideal for this as they are a series of compounds with a basic di-peptide frame that had different substituents, side chains and chirality. They also have IC_{50} data against rat μ-calpain that can be compared with the docking data. The docking data obtained from the study of the eighteen compounds was shown to be consistent with the IC_{50} data. It was shown that compounds such as **3.10** with good inhibition of calpain docked into the model with all the required parameters to suggest it was a good inhibitor. Conversely, compounds that had poor IC_{50}'s such as **3.22** showed that they could not dock in the required manner by having a poor Glide score, less than the required three H-bonds, a warhead further than 5Å from Cys115, and/or an excessive number of 'ugly' internal contacts. The important point here is that docking of the compounds into the Glide model generated data that predicted, with reasonable accuracy, the likelihood that a compound would inhibit calpain *in vitro*. Now that we had a working viable model we could proceed with docking of our in house compounds and making a contribution as to what compounds were likely to be inhibitors and therefore worthy of synthesis.

Next, a series of *N*-heterocyclic compounds based on the lead compound **3.10** were suggested as possible candidates by our synthetic chemists following from the work of other groups showing that addition of *N*-terminal capping groups could improve selectivity and membrane permeability. Specifically we have investigated the potential of different *N*-terminal capping groups to interact with the P_3 position in the enzyme.

The Glide docking studies showed that most of the compounds (**3.26**-**3.35**) could dock in the required manner. However, **3.26** and **3.33** docked into the model in a previously unseen

pose. They still had all the requirements of good inhibitors but the three essential H-bonds were in a different position on the ligand. Compound **3.34** did not dock with Glide but did dock appropriately when the InducedFit Protocol was implemented.

Based on their 'unique' docking poses, compounds **3.26** and **3.33** were synthesized and tested against both μ- and m-calpain along with all the other *N*-heterocyclic compounds. All the compounds were shown to be good or very good inhibitors of μ- and m-calpain. Compound **3.26** turned out to be the most potent against m-calpain (IC_{50} of 25 nM) and was ten times more selective for m-calpain over μ-calpain and compound **3.33** was the most potent against m-calpain (IC_{50} of 150 nM) but was not selective. Docking had predicted that most of these compounds would be good inhibitors of calpain and that the 'unique' docking of compounds **3.26** and **3.33** suggested that these compounds were worthy candidates for synthesis.

The series of diazo- and triazene-dipeptide aldehydes **3.36a-d**, **3.37a**,**b**, **3.38a**,**b** prepared predominantly as the (*E*)- isomers are all highly potent against m-calpain, with the most potent **3.36b** [5.4(*E*)/1(*Z*)] having an IC50 value of 35 nM. Photoisomerism of the diazo inhibitors **3.36a-c** and **3.37a** gave samples enriched in the (*Z*)-isomer that proved to be significantly less active. SAR data suggests that an *N*-terminal diazo group (**3.36a-d**) is favoured over a triazene (**3.38a**,**b**). Furthermore 4-substitution of the *N*-terminal diazo group (**3.36a**,**b**) is favoured over 3-substitution (**3.36c**,**d**). The triazene of **3.38a** imparts improved water solubility for *in vivo* studies and was shown to arrest the development of calpain-induced cataract formation in sheep lens culture.

Molecular modeling predicts that these compounds all bind in an extended β-strand conformation, as defined by three key hydrogen bonds to Gly208 and Gly271. This is consistent with published crystal structures of μ-calpain with the inhibitor bound demonstrating the validity of the modelling methodology.[5,6,20] The reactive carbonyl in each case is located in close proximity to the active site cysteine, as would be required for mechanism-based inhibition. As a consequence, the *N*-terminal group of the (*E*)-diazo dipeptide aldehydes **3.36a-d** and **3.37a**,**b** extends deep into the S_3 binding pocket.

A calpain inhibitor, the activity of which can be influenced by irradiation, offers some potential as a means to actively control cataracts and this is a first step toward such a goal.

3.6 References

1. Donker, I. O., A Survey of Calpain Inhibitors. *Curr. Med. Chem.* **2000,** *7* 1171-1188.
2. Maki, M.; Takano, E.; Mori, H.; Sato, A.; Murachi, T.; Hatanaka, M., All four internally repetitive domains of pig calpastatin possess inhibitory activities against calpains I and II. *FEBS Letters* **1987,** *223*, 174-180.
3. Todd, B.; Moore, D.; Deivanayagam, C. C. S.; Lin, G.; Chattopadhyay, D.; Maki, M.; Wang, K. K. W.; Narayana, S. V. L., A Structural Model for the Inhibition of Calpain by Calpastatin: Crystal Structures of the Native Domain VI of Calpain and its Complexes with Calpastatin Peptide and a Small Molecule Inhibitor. *J. Mol. Biol.* **2003,** *328*, 131-146.
4. Kiss, R.; Kovacs, D.; Tompa, P.; Perczel, A., Local Structural Preferences of Calpastatin, the Intrinsically Unstructured Protein Inhibitor of Calpain. *Biochemistry* **2008,** *47*, 6936–6945.
5. Moldoveanu, T.; Campbell, R. L.; Cuerrier, D.; Davies, P. L., Crystal Structures of Calpain-E64 and-Leupeptin Inhibitor Complexes Reveal Mobile Loops Gating the Active Site. *J. Mol. Biol.* **2004,** *343* 1313.
6. Cuerrier, D.; Moldoveanu, T.; Inoue, J.; Davies, P. L.; Campbell, R. L., Calpain Inhibition by Alpha-Ketoamide and Cyclic Hemiacetal Inhibitors Revealed by X-ray Crystallography. *Biochemistry* **2006,** *45*, 7446-7452.
7. Qian, J.; Cuerrier, D.; Davies, P. L.; Li, Z.; Powers, J. C.; Campbell, R. L., Cocrystal Structures of Primed Side-Extending α-Ketoamide Inhibitors Reveal Novel Calpain-Inhibitor Aromatic Interactions. *J. Med. Chem.* **2008,** *51*, 5264–5270.

8. Tsujinaka, T.; Y, Y. K.; Kambayashi, J.; Sakon, M.; Higuchi, N.; Tanaka, T.; Mori, T., Synthesis of a new cell penetrating calpain inhibitor (calpeptin). *Biochemical and biophysical research communications* **1988,** *153*, 1201-8.

9. Mehdi, S.; Angelastro, M. R.; Wiseman, J. S.; Bey, P., Inhibition of the proteolysis of rat erythrocyte membrane proteins by a synthetic inhibitor of calpain. *Biochem. Biophys. Res. Comm.* **1988,** *157*, 1117-23.

10. Sasaki, T.; Kishi, M.; Saito, M.; Tanaka, T.; Higuchi, N.; Kominami, E.; Katunuma, N.; Murachi, T., Inhibitory effect of di- and tripeptidyl aldehydes on calpains and cathepsins. *Journal of enzyme inhibition* **1990,** *3*, 195-201.

11. Inoue, J.; Nakamura, M.; Cui, Y.; Sakai, Y.; Sakai, O.; Hill, J. R.; Wang, K. K. W.; Yuen, P., Structure-Activity Relationship and Drug Profile of N-(4-Fluorophenylsulfonyl)- L valyl-L-leucinal (SJA6017) as a Potent Calpain Inhibitor. *J. Med. Chem.* **2003,** *46*, 868-871.

12. Thompson, V. F.; Saldana, S.; Cong, J.; Goll, D. E., A BODIPY fluorescent microplate assay for measuring activity of calpains and other proteases. *Anal. Biochem.* **2000,** *279*, 170–178.

13. Fairlie, D. P.; Tyndall, J. D. A.; Reid, R. C.; Wong, A. K.; Abbenante, G.; Scanlon, M. J.; March, D. R.; Bergman, D. A.; Chai, C. L. L.; Burkett, B. A., Conformational Selection of Inhibitors and Substrates by Proteolytic Enzymes: Implications for Drug Design and Polypeptide Processing. *J. Med. Chem.* **2000,** *43* 1271-1281.

14. *MacroModel*, version 9.1; Schrödinger: LLC, New York, NY, 2005.

15. *MacroModel XCluster*, version 9.1; Schrödinger: LLC, New York, NY, 2005.

16. Assay of irradiated **8a**, **8c**, and **10a** enriched in the (*Z*)-isomer were carried out in dim light conditions, that is, glassware wrapped in foil and with the lights turned off.

17. The calpain inhibitor SJA6017 **1** was chosen as a literature standard to validate our assay procedure and for comparison of inhibitor potency.

18. *Schrödinger Suite 2006 Induced Fit Docking Protocol*, Glide version 4.0, Prime version 1.5; Schrödinger: LLC, New York, NY, 2005.

19. The X-ray crystal structure for rat mu-calpain was used for the docking studies as the m-calpain X-ray crystal structure was in an inactivated form. Structures for sheep m- or mu-calpain have not been published; however analysis indicates 94% homology between sheep and rat m-calpain.

20. Li, Q.; Hanzlik, R. P.; Weaver, R. F.; Schoenbrunn, E., Molecular Mode of Action of a Covalently Inhibiting Peptidomimetic on the Human Calpain Protease Core. *Biochemistry* **2006,** *45*, 701-708.

21. For example, the (*E*)-isomer of **8a** is calculated to be 40-fold more active than the (*Z*)-isomer.

4 Molecular modelling of cyclic inhibitors

4.1 Introduction

Cyclic inhibitors can have an advantage over an acyclic counterpart. Cyclization can preorganize functionality on a ligand to dictate conformations that the target enzyme readily recognizes and in so doing can lower the entropic barrier for inhibitor-enzyme complex formation.[1-5] Acyclic inhibitors, such as small peptide like compounds, contain a number of rotatable bonds. The more rotatable bonds a ligand has the more conformationally flexible the ligand is likely to be, and because a ligand binds to an enzyme in a specific conformation it can be desirable to minimise this entropic flexibility. As we have already discussed native polypeptide substrates of protease enzymes are often folded into well defined conformations which are almost exclusively of the β-strand variety.[6-8] Locking functionality of a ligand into such a conformation can be achieved by cyclization which reduces conformational flexibility, with a consequent increase in binding affinity.[1]

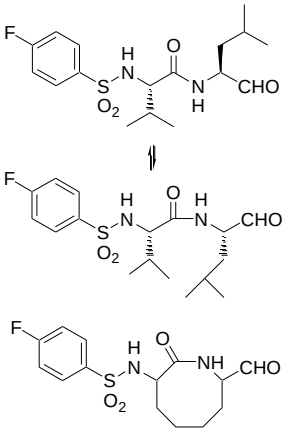
4.2 First generation cyclic analogues: modelling studies of 8-membered cyclic analogues of SJA-6017

One of the most potent inhibitors discussed so far is *N*-(4-fluorophenylsulfonyl)-L-valyl-L-leucinal (SJA-6017) (compound **3.10** in **Figure 4.1**).[9, 10] This compound displays potent inhibition of calcium activated m-calpain (IC_{50} = 80 nM)[9] and has been shown to be effective against selenite induced cataract formation in rats[11] and in cataractic porcine lenses.[12] In addition it has low toxicity, it is specific for cysteine proteases, but suffers from low

permeability into the cell and it binds indiscriminately and readily to proteins. *In vitro* studies with human plasma have shown that 95.4% of **3.10** at a concentration of 10 μM is bound to the plasma proteins.[10] Therefore, it is necessary to make structural modifications to this compound without compromising its potency or specificity.

Our first studies, before coordinates of the μ-capain construct were known, focused on the cyclic analogues **4.1** of **3.10** with the leucinal and valyl side chains modified and linked to form an 8-membered ring.[13-15] This structural modification restricts the conformational flexibility of the amide linkage. The non natural environment of the amide linkage will alter its proteolytic stability as proteases may be less able to recognize it. Furthermore, the modification will change the hydrophilicity and thereby the ability to cross membranes, a major problem of the lead compound **3.10**. The restricted conformational space brought about by the 8-membered ring may also decrease the unspecific binding to other proteins. The results of the activity tests of **4.1**, having limited conformational flexibility, is intended to allow us to investigate further the requirements for molecular recognition in the active site of calpain. In general, most proteases are considered to preferably cleave amide bonds that are part of a β-strand secondary conformation.[2, 6] The synthetic route for these compounds follows that by Creighton et al.[13]

3.10



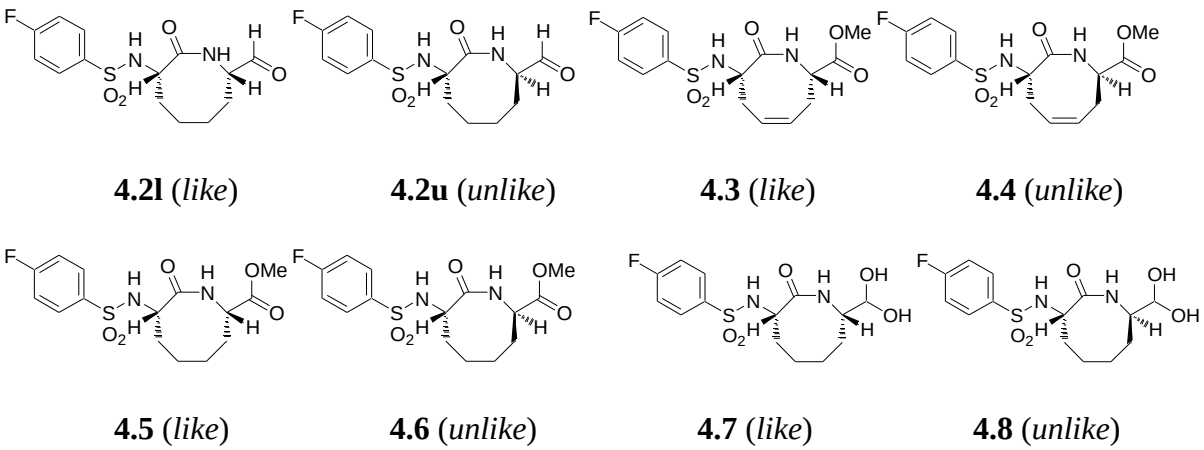
4.1

Figure 4.1: The known calpain inhibitor SJA0617 (**3.10**) resembles the cyclic analogue **4.1** when adopting a *cis* peptide bond conformation.

The first generation cyclic analogues of SJA6017 (**3.10**) that were prepared in our group were therefore synthesized before we had suitable models of the enzyme. The leucinal and valyl side chains were modified and linked to form an eight-membered ring thereby restricting conformational flexibility of the intramolecular amide bond. It was hoped that such a cyclic analogue would help overcome problems of SJA6017 and be a drug candidate without compromising potency or specificity. The problems associated with SJA6017 include being too hydrophilic to cross membranes (low cell permeability) and binding indiscriminately and readily to other proteins resulting in low bioavailability.[10]

Compounds **4.3**-**4.8** (**Figure 4.2**) were tested for calpain inhibition and shown to be inactive and now, with appropriate enzyme model coordinates, subsequent modelling, which we now report, demonstrated why they were inactive. The fact that the equilibrium mixture of diols **4.7** and **4.8** and aldehydes **4.2*l*** and **4.2*u*** were inactive is particularly significant as it showed the importance of the β-strand in calpain inhibitor design.

Modeling studies using MacroModel[16] were used to establish the global minima and ensemble of low energy conformations of the compounds **4.3**-**4.8, 4.2*l*** and **4.2*u*** in simulated water using **Protocol 3** in the **Appendix**. The ensembles of generated structures were clustered using the program X-cluster[17] and the clusters within 12 kJ/mol window of the global minima were analysed. The clusters represent the backbone motifs for these cyclic conformationally restricted structures. It is worth noting that the *unlike* diastereomers **4.4**, **4.6**, **4.8** and **4.2*u*** contain one amino acid of unnatural configuration. Also none of the diastereomers can adopt typical peptide secondary structural motifs such as α-helical, β-strand or β-turns. The global minima conformations for **4.3** and **4.4** are shown in **Figure 4.3** and **Figure 4.4** as examples of major backbone conformations of *like* and *unlike* diastereomers. The Boltzmann weighted calculations of the ensembles of structures gave for each diastereomer the potential energy, distances between the chiral methine protons, the number of hydrogen bonds, and provided for the determination of the ratio of the number of hydrogen bonds for the diastereomers (**Table 4.1**).



4.2l (*like*) **4.2u** (*unlike*) **4.3** (*like*) **4.4** (*unlike*)

4.5 (*like*) **4.6** (*unlike*) **4.7** (*like*) **4.8** (*unlike*)

Figure 4.2 Compounds modelled and tested:

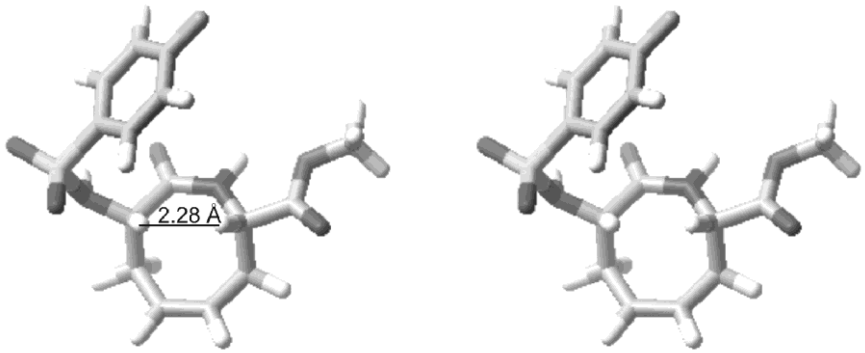


Figure 4.3: Stereo view of the global minimum conformation of **4.3**.

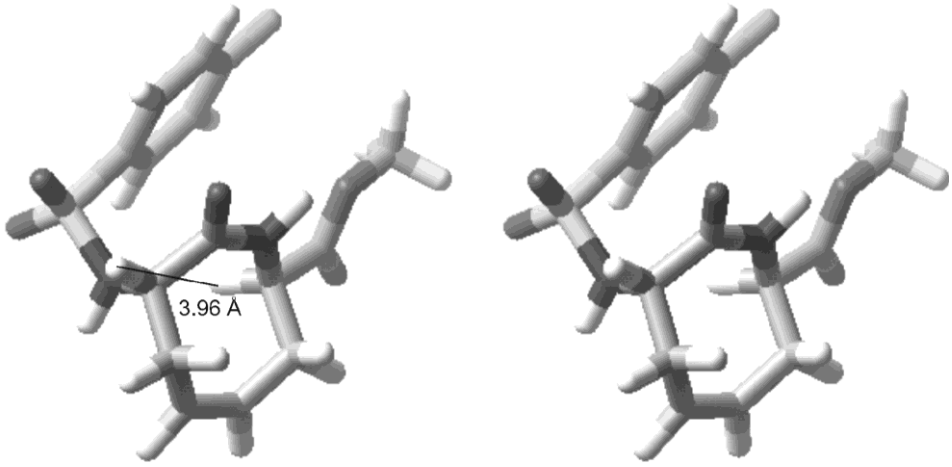


Figure 4.4: Stereo view of the global minimum conformation of **4.4**.

Table 4.1: Modelling data for the lowest energy conformation of **4.2*l***, **4.2*u*** and **4.3**-**4.8**.

	4.3	**4.4**	**4.5**	**4.6**	**4.7**	**4.8**	**4.2 *l***	**4.2 *u***
a	96.5	104.4	-3.5	7.6	-130.4	-129.6	-12.9	-5.8
b	7.9		11.1		0.8		7.1	
c	2.3	4.0	2.1	3.9	2.0	3.9	2.1	3.7
d	2.3	4.0	2.1	3.9	2.1	3.8	2.1	3.8
e	2.0	1.0	2.0	1.0	2.0	2.0	2.0	0.9
f	1.7:1		2.0:1		0.9:1		1.7:1	
g	2.1:1		2.0:1		1.0:1		2.1:1	

a: Boltzmann weighted average ensemble energy [kJ/mol]*; b:* energy difference between the ensembles of like and unlike diastereomers [kJ/mol]; *c*: average distance between the chiral methine protons of the ensembles [Å]; *d*: Boltzmann weighted distance between the chiral methine protons of the ensembles [Å]; *e*: Boltzmann weighted average number of H-bonds per conformer in the ensembles; *f*: average ratio of the number of H-bonds of the *like* and *unlike* diastereomers; *g*: Boltzmann weighted average of the *like*:*unlike* ratio of H-bonds.

Of the hydrated aldehydes **4.7** and **4.8** the major diastereomer is **4.7**, which is calculated to be the more thermodynamically favoured, but by only 0.8 kJ/mol. The higher thermodynamic stability of the *like* diastereomers **4.3**, **4.5**, and **4.2*l*** is explained by the modeling results (see **Table 4.1**). The Boltzmann weighted energy of the ensemble of conformers of the *like* diastereomers is in all cases lower for the corresponding *unlike* diastereomers **4.4**, **4.6**, and **4.2*u***. This observation correlates with the higher average and the higher Boltzmann weighted average number of internal hydrogen bonds for the *like* diastereomers (see **Table 4.1**).

It is notable that for the hydrated aldehydes **4.7** and **4.8** the energy difference is significantly lower than between all the other pairs of diastereomers. Also the difference in the average and the Boltzmann weighted average number of internal hydrogen bonds between the diastereomers is small (see **Table 4.1**). This supports the importance of intramolecular hydrogen bonding in differentiating the energy between the diastereomers.

The preferential formation of **4.7** over **4.8** can be explained most convincingly by the energy difference of the precursors **4.2*l*** and **4.2*u*** that differ more in energy than the diols.

The relative configuration of the diastereomers was determined from molecular modelling and NMR studies where the distances between the methine protons correlate with NOE measurements. For example an appropriately weighted distance of 2.28 Å between the chiral methine protons of **4.3** (**Figure 4.3**) is consistent with the observed NOE for the *like* diastereomer. While for the *unlike* diastereomer **4.4** (**Figure 4.4**) the weighted distance was 3.96 Å and no NOE could be detected. Similar weighted distances and NOE observations of

the chiral methine protons were apparent for all *like*/*unlike* pairs, confirming the relative stereochemistry of the chiral carbons.

The absence of calpain inhibition by all of the 8-membered cyclic compounds, but especially by diols **4.7** and **4.8**, which will be in equilibrium with the aldehydes **4.2*l*** and **4.2*u***, is significant. The compounds are not able to adopt the typical peptide secondary structural motifs *β*-strand, *β*-turns, or *α*-helix. **Figure 4.5** shows compound **3.10**, which can form the 'zigzag' pattern of a *β*-strand along the backbone, superimposed with compound **4.2*u***, which cannot form this type of conformation. This is because the 8-membered ring forces the backbone of the dipeptide to form a *cis* conformation at the peptide bond rather than the *trans* conformation required for *β*-strand formation. These results are in accordance with the assumption that a *β*-strand structure of a compound is required for molecular recognition by calpain as shown for many known protease inhibitors and substrates.[2, 6]

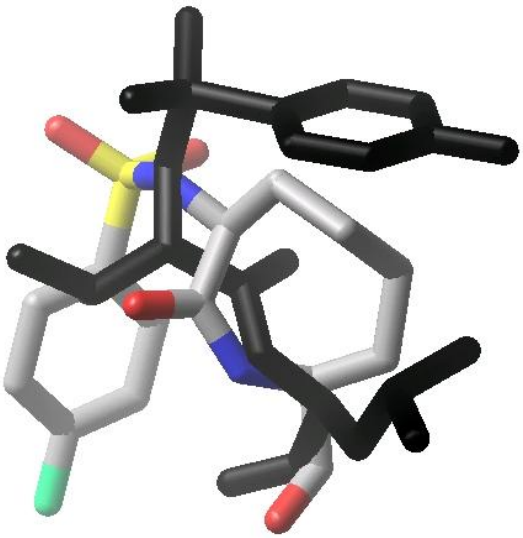


Figure 4.5. The lowest energy conformer of SJA-6017 (**3.10**) (black) showing a *β*-strand conformation superimposed with compound **4.2*u***, which inherently cannot form a *β*-strand.

In summary, a series of conformationally restricted 8-membered cyclic analogues of **3.10** were synthesized, tested, and modelled. The higher calculated thermodynamic stability of the *like* diastereomers **4.3, 4.5**, and **4.2*l*** correlates with the prevalence of internal hydrogen bonds. The compounds are not able to adopt typical peptide secondary structural motifs. No inhibition of any of the compounds in this study, especially by diols **4.7** and **4.8**, which will be in equilibrium with the aldehydes **4.2*l*** and **4.2*u***, could be shown. These results are in agreement with the putative importance of a β-strand structure for inhibitors and substrates of calpain, a structural feature exhibited for several known protease inhibitors. This gives further information on the requirements for molecular recognition in the active site of calpain.

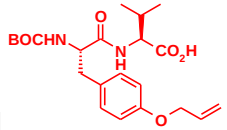
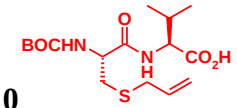
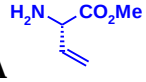
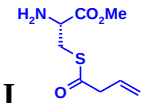
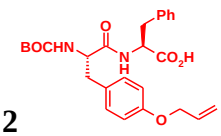
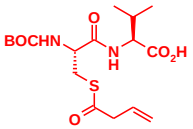
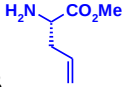
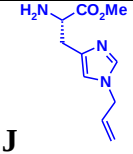
4.3 Second generation cyclic analogues: modelling studies of ca 17 membered macrocyclic ring analogues of SJA-6017

We then attempted to design small conformationally constrained macromolecules that form a β-strand type conformation so that the amino acid segment of the molecule positions functional groups appropriately for binding with the calpain enzyme. This second generation of cyclic inhibitors have been based on Fairlie's[6] analysis of the importance of a β-strand for bioactivty. He reported in an important publication that nearly all known substrates of known proteases bind in an extended β-strand conformation.[2] The aim of my work has been directed to constraining an inhibitor in such a bioactive conformation, thus, lowering the entropic barrier for binding of the inhibitor to the enzyme.

Fairlie[6] reported that incorporating a tripeptide into a ca 17 membered ring favours the tripeptide in an extended β-strand conformation.[2, 6-8] To investigate if we could conformationally constrain our lead compound (SJA6017) as a macrocycle, in a β-strand, a

library of possible candidates (structures) were devised by Dr Steve Aitken, then a PhD student. For inclusion of a candidate in the virtual library we first required that a reasonable synthetic route be available. The virtual library is constructed from an 18x16 grid matrix shown in **Table 4.2** to yield 288 possible tripeptide macrocyclic structures. The proposed synthetic route to generating each macrocycle is as follows: An *N*-BOC-allyl-amino acid with an olefin in the side chain (left columns **Table 4.2**) is coupled with an allyl-amino acid methyl ester also with an olefin in the side chain (right columns **Table 4.2**) to yield the corresponding diene. Ring closing metathesis can be used to generate the macrocycles, a method proven by Dr Sigeru Miyamoto's study of compounds **4.2*u***, **4.2*l***, and **4.3**-**4.8**.

These 288 possible core structures could be modified to give different warheads and protecting groups. For the initial conformational search the warhead for each of the 288 structures was an aldehyde and the protecting group was 4-fluorobenzyl sulphonamide. For example, *N*-BOC-allyl-amino acid **1** coupled with allyl-amino acid methyl ester **A** would give structure **X** (**Figure 4.5**). If **X** was modified to be an aldehyde protected with 4-fluorobenzyl sulfonamide it would give structure **A1** in **Figure 4.5**. A conformational search of each structure generated ensembles of the low energy conformers described in **Protocol 3** in the **Appendix**.

***N*-BOC-allyl-amino acid**		**Allyl-amino acid methyl ester**	
 **1**	 **10**	 **A**	 **I**
 **2**	 **11**	 **B**	 **J**

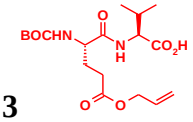
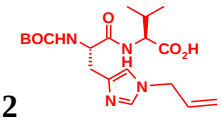
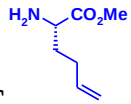
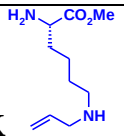
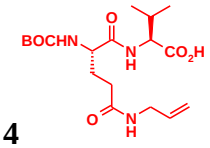
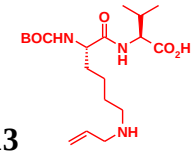
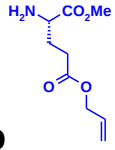
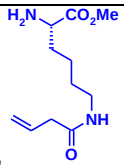
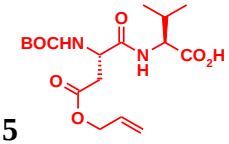
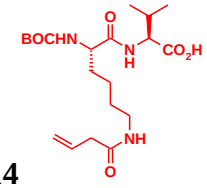
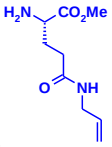
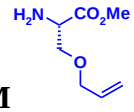
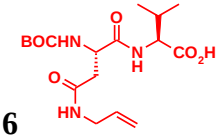
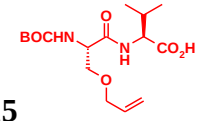
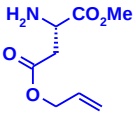
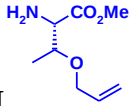
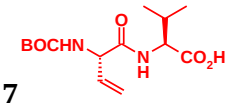
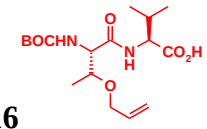
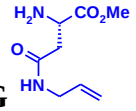
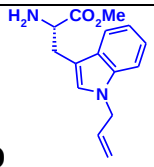
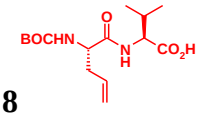
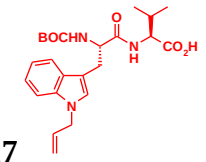
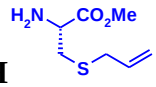
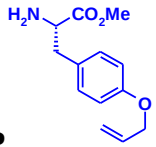
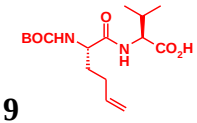
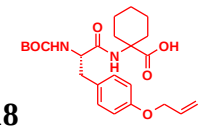
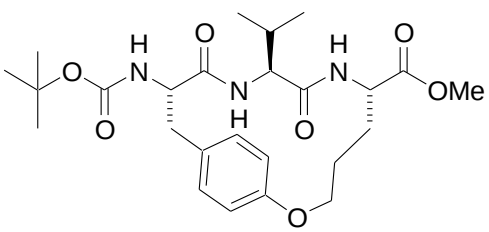
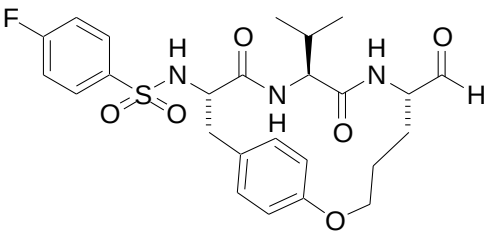
3 	**12** 	**C** 	**K** 
4 	**13** 	**D** 	**L** 
5 	**14** 	**E** 	**M** 
6 	**15** 	**F** 	**N** 
7 	**16** 	**G** 	**O** 
8 	**17** 	**H** 	**P** 
9 	**18** 		

Table 4.2 Acids **1**-**18** coupled with esters **A**-**P** and closed with ring closing metathesis to give *in silico* library of 288 macrocyclic tripeptides. These were modified *in silico* to give the equivalent 4-fluorobenzyl sulphonamide protected aldehydes to give compounds labelled **A1**, **A2**, **A3**...**P16**, **P17**, **P18**.



X



A1

Figure 4.5 X is an example of the structure when acid **1** is coupled with ester **A** and closed with ring closing metathesis. **A1** is the 4-fluorobenzyl sulphonamide protected aldehyde of **X** used for the conformational search.

4.3.1 Docking of macrocyclic analogues of SJA6017 using Glide

We had not performed any docking experiments of the macrocycles before and did not know if the Glide model would accommodate these comparatively large compounds (compared with the smaller di-peptides we had previously docked). In order to test if the Glide model would work we took the first ten compounds, **A1**-**A4**, **B1**-**B4**, **C1** and **C2** (**Figure 4.6**), in the 288 compound library and docked them using Glide with the parameters shown in **Protocol 4** in the **Appendix**. These ten compounds display macrocyclic ring sizes from 14-18 which gave a range of sizes to test.

The following is Glide docking data of ten selected compounds from the matrix: The lowest energy conformer of compounds **A1**-**A4**, **B1**-**B4**, **C1** and **C2** (**Figure 4.6**) were docked into the μ-calpain construct (1KXR) model using Glide (**Protocol 4** in the **Appendix**).

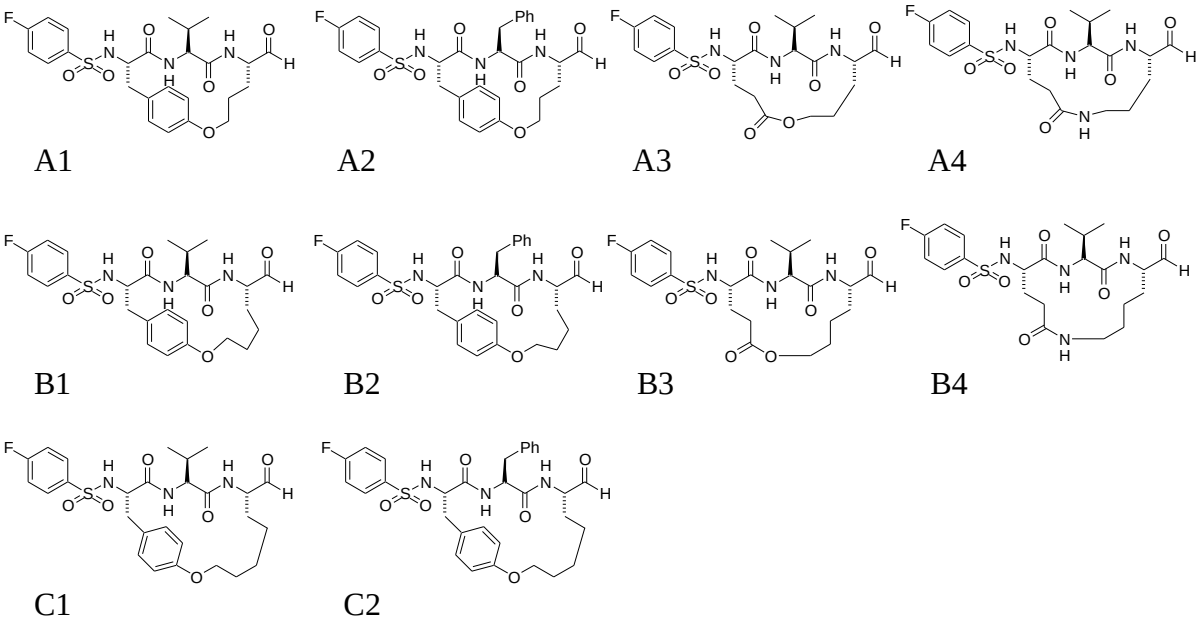


Figure 4.6; Chemdraw structures of compounds **A1**-**A4**, **B1**-**B4**, **C1** and **C2**.

Compounds **A1**, **A4**, and **C1** docked into the active site with the essential criteria of the three H-bonds displaying a *β*-strand, close electrophile-nucleophile proximity, and low Glide and E-model scores (see **Table 4.3** and **Figure 4.7)**. The capping group of each is oriented so that it sits over Gly 208.

Compound **A1** has been subsequently synthesized and tested in house and has a moderate IC_{50} of 3710nM against ovine m-calpain. Compounds **A1** and **A4** both have some 'ugly' internal contacts, a possible reflection on compound **A1**'s less than flattering IC_{50}. 'Ugly' internal contacts mean that the compound has had to be put in a slightly unfavourable conformation in order to fit into the active site. In a real system, movement in the active site, which may or may not be better for enzyme-inhibitor binding, could allow the compound to bind in a slightly more favourable conformation.

Table 4.3: Docking data for best pose[a] (out of possible 10) and inhibitory concentrations (IC_{50}) for compounds **A1**-**A4**, **B1**-**B4**, **C1** and **C2**

Compound	Glide Score [b]	emodel Score [c]	Essential H bonds	Warhead Distance Å	Internal contacts			IC_{50} (nM) m-calpain
					Good	Bad	Ugly	
A1	-5.1	-54.9	3	4.4	280	22	2	3710
A2	-4.3	-56.4	0	>5	295	26	0	
A3	-4.7	-40.8	0	>5	282	13	2	
A4	-4.3	-48.8	3	4.7	259	16	3	
B1	-5.4	-43.7	0	>5	296	15	1	280
B2	-5.9	-49.4	0	>5	299	23	1	
B3	-4.4	-42.9	0	>5	284	11	1	
B4	-4.4	-43.8	0	>5	287	16	0	
C1	-4.4	-52.2	3	4.2	327	16	0	
C2	-5.5	-60.3	0	>5	330	18	0	

[a] Best pose chosen by criteria in order of importance; 1 - presence of three essential hydrogen bonds, 2 - warhead within 5Å of nucleophilic Cys115, 3 - low Glide score/emodel score, 4 - lowest number of internal ugly contacts.

[b] GlideScore[18] is a scoring function based on ChemScore[19] which is designed to estimate free energy of binding for protein–ligand complex. The function uses simple contact terms to estimate lipophilic and metal–ligand binding contributions, a simple explicit form for hydrogen bonds and a term which penalises flexibility.

[c] Emodel[18] is a model energy score that combines energy grid score, binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy.

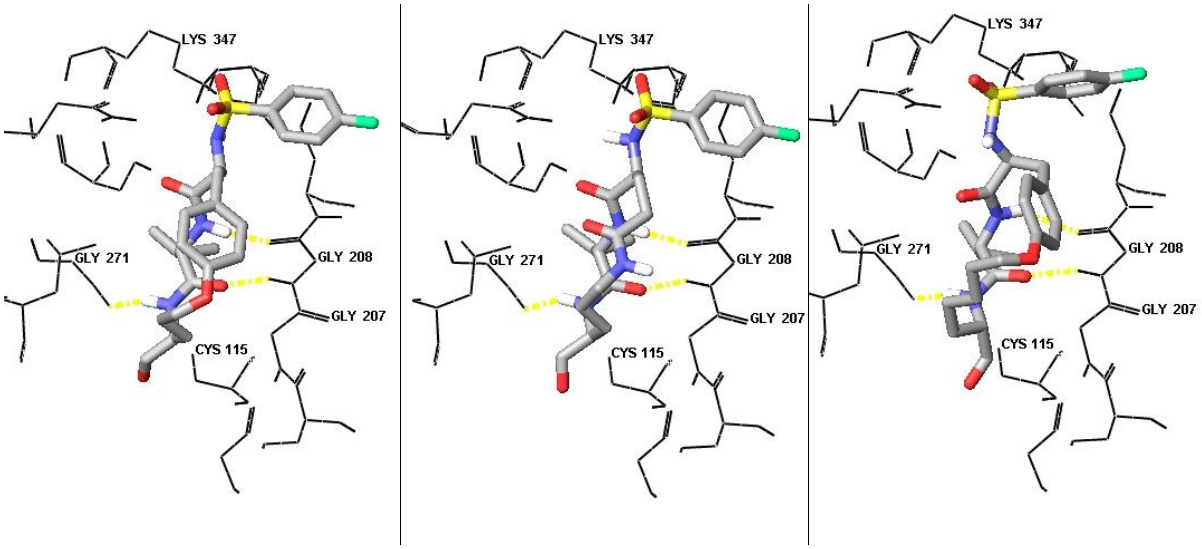


Figure 4.7: Best docked pose of compound **A1** (left) compound **A4 (**middle), and compound **C1 (**right). All docked with Glide into μ-calpain model.

The other seven compounds in this docking experiment did not dock into the enzyme model with the proposed essential criteria as can be seen in **Figure 4.8** (showing only docked poses of **A2**, **A3**, and **B3**) and **Table 4.3.** It was also noticed that some of these docked compounds were not in a β-strand conformation. Of these seven compounds only **A2** was found to have its starting conformation in a β-strand.

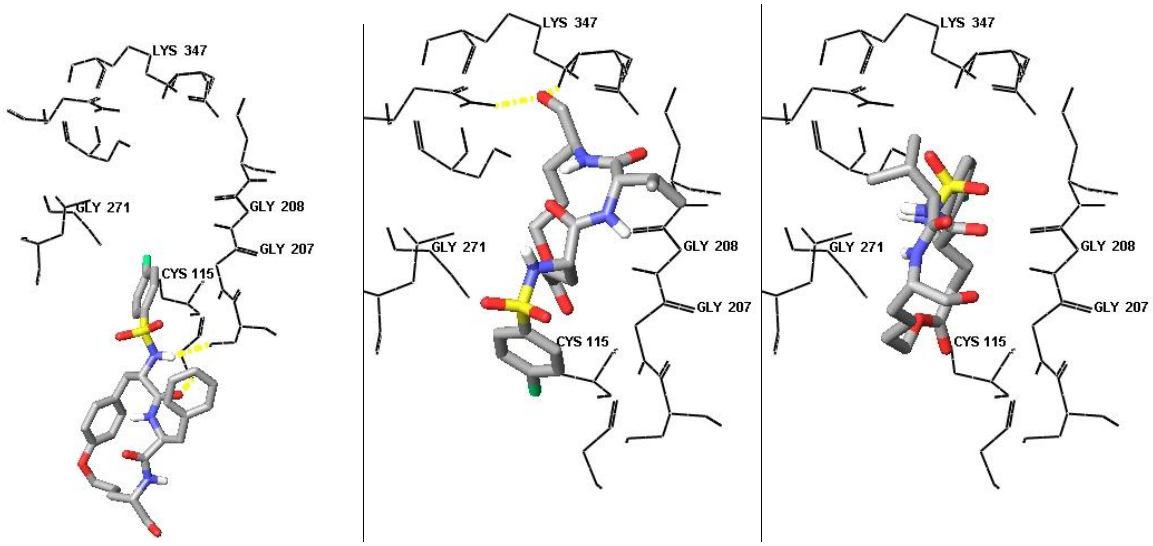


Figure 4.8: Best docked pose of compound **A2** (left), compound **A3 (**middle), and compound **B3 (**right). All docked with Glide into μ-calpain model.

Compound **A2** displays a β-strand conformation but none of the docked poses of **A2** showed that the compound could form the three essential H-bonds. **A2** differs from **A1** only at the P2 position where a phenylalanine replaces a valine. The bulkier phenylalanine side chain of compound **A2** may not fit into the deep hydrophobic pocket of the enzyme model where P2 side chains are known to bind.[20-22]

After this initial docking experiment was performed and it was noticed that some of the docked compounds were not in a β-strand and a closer look at the ensemble of low energy macrocycles was made. Many of the low energy conformers of the 288 compounds were found to not be in the intended β-strand and were in fact displaying a type of turn, and in a few cases they were displaying a conformation in between a β-strand and a γ-turn. I have called this structure a 'twist' conformer. The turns are actually a form of a γ-turn which is a rare conformational feature in natural peptides. Examples of a macrocycle in a β-strand and a turn are shown in **Figure 4.9**.

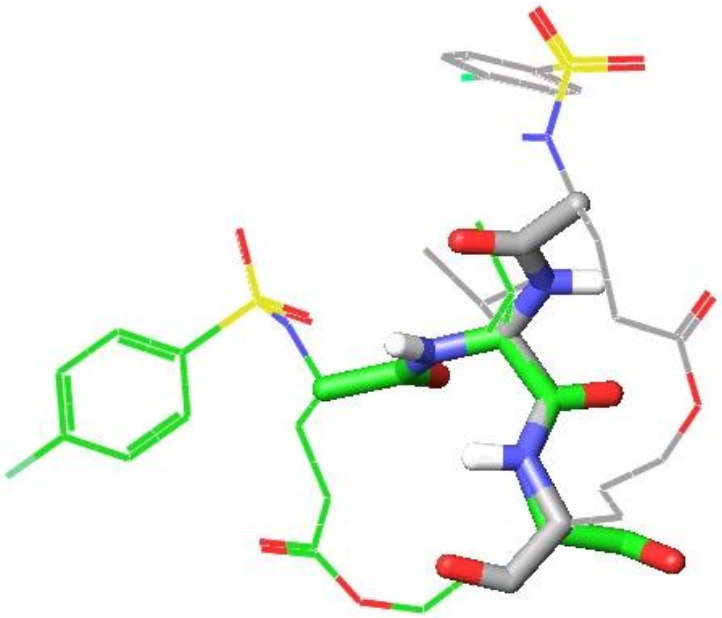


Figure 4.9. Macrocycle **A3** in a low energy *β*-strand conformation (grey carbons) superimposed with a low energy conformation of **A3** in a turn type conformation (green carbons).

The ensembles of low energy conformers of the six structures (compounds **A3**, **B1**, **B2**, **B3**, **B4**, and **C2**) that did not have a *β*-strand conformation as the lowest energy conformer (used as the starting conformer for initial docking) were examined to determine (see discussion below on appropriate parameters to describe a *β*-strand) the lowest energy conformer that displayed a *β*-strand, if any. These low energy conformers that displayed a *β*-strand were then docked with the same parameters as before to determine if they could bind into the active site with the essential criteria when their starting conformer was in a *β*-strand.

Compounds **B1**, **B2**, and **B3** when in a *β*-strand starting conformation were shown to dock into the active site with all or most of the essential criteria intact (**Table 4.4**).

Table 4.4: Docking data for best pose[a] (out of possible 10) and inhibitory concentrations (IC_{50}) for compounds **B1**, **B2**, and **B3** when starting conformer was the lowest energy conformer in a *β*-strand.

Compound	Glide Score [b]	emodel Score [c]	Essential H bonds	Warhead Distance Å	Internal contacts			IC_{50} (nM) m-calpain
					Good	Bad	Ugly	
B1	-4.4	-52.1	3	3.5	284	19	4	280
B2	-5.5	-60.1	2	3.9	310	19	1	
B3	-6.6	-49.3	3	3.5	254	11	0	

[a] Best pose chosen by criteria in order of importance; 1 - presence of three essential hydrogen bonds, 2 - warhead within 5Å of nucleophilic Cys115, 3 - low Glide score/emodel score, 4 - lowest number of internal ugly contacts.

[b] GlideScore[18] is a scoring function based on ChemScore[19] which is designed to estimate free energy of binding for protein–ligand complex. The function uses simple contact terms to estimate lipophilic and metal–ligand binding contributions, a simple explicit form for hydrogen bonds and a term which penalises flexibility.

[c] Emodel[18] is a model energy score that combines energy grid score, binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy.

When compound **B1** was docked, using the lowest energy conformer in a *β*-strand, into the enzyme model all the essential criteria was displayed (**Figure 4.10** left), however, four 'ugly' internal contacts were identified.

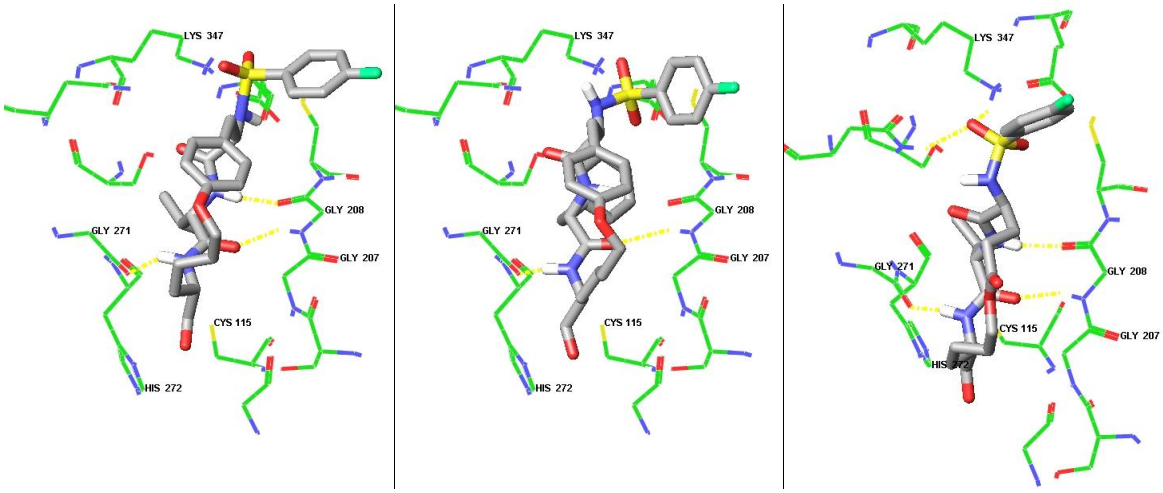


Figure 4.10: Best docked pose of compound **B1** (left), compound **B2 (**middle), and compound **B3 (**right). All docked with Glide into µ-calpain model.

Compound **B2** docked using Glide with all essential elements of the required criteria except only two of the three essential H-bonds were observed (**Figure 4.10** middle). The third

H-bond was absent but the atoms usually involved in this H-bond were relatively close to being able to form such a bond. **B2**, like **A2**, has a phenylalanine in the P_2 position which requires more space in the S_2 subsite due to being bulkier than the valine side chain of compounds like **B1**. This fact may be why **B2** is docked somewhat higher out of the pocket which consequently causes the backbone NH to be further away from the backbone oxygen of Gly208 and so no H-bond was observed. It also had one 'ugly' internal contact.

Compound **B3** was shown to dock in the required manner when the lowest energy conformer in a *β*-strand was used as the starting conformer in the docking experiment (**Table 4.4** and **Figure 4.10** right). It had a low Glide score of -6.6 and showed no 'ugly' internal contacts.

In contrast, compounds **A3**, **B4**, and **C2** did not dock in the required manner even when docking was performed using their lowest energy conformers in a *β*-strand.

Overall, the Glide docking study showed that compounds **A1**, **A4, B1**, **B3**, and **C1** could dock with Glide into the active site of the Glide calpain model with the three critical H-bonds intact, with a warhead to nucleophile distance of less than 5Å, and with low Glide and Emodel scores. This occurred only when the lowest energy conformer in a *β*-strand was used as the starting conformer in Glide. Of these five compounds only **B3** and **C1** had no 'ugly' internal contacts. The other three compounds in this group had docking poses with a low number of 'ugly' internal contacts.

The rigidity of the Glide enzyme model (a limitation of Glide docking) was hypothesized to be a possible factor as to why some of the compounds with larger P2 side chains, namely compounds **A2**, **B2**, and **C2**, could not fit into the active site in the appropriate way. It was

also possible that the side chain of Lys347 was hindering the docking of some compounds as its proximity was in the region where the ligand capping groups could possibly dock. The long side chain of Lys347 has the ability to reposition itself to form H-bonds to atoms of the ligand capping groups.

For the above reasons it was deemed necessary and best to use Schrodinger's InducedFit Protocol for subsequent docking. The InducedFit Protocol uses a program script that combines the docking program Glide with the protein structure prediction suite called Prime with the aim of simulating movement of the active site upon ligand binding. It was hoped that movement of the active site would also help some ligands 'relax' so that some of the 'ugly' internal contacts seen in the pure Glide docking would not form. The side chain of Lys347 could also be temporarily removed during the initial docking phase of the InducedFit Protocol to allow an easier fit for the macrocycles and subsequent movement of the flexible side chain of Lys347.

4.3.2 Docking of macrocyclic analogues of SJA6017 using InducedFit

The compounds **A1**-**A4**, **B1**-**B4**, **C1** and **C2** (**Figure 4.6**) used in the previous section for Glide docking were subsequently docked into the enzyme model using the InducedFit Protocol (**Protocol 5** in the **Appendix**).

Table 4.5: Docking data for best pose[a] (out of possible 20) and inhibitory concentrations (IC_{50}) for compounds **A1**-**A4**, **B1**-**B4**, **C1** and **C2** when InducedFit docking was used.

Compound	Glide Score [b]	emodel Score [c]	Essential H bonds	Warhead Distance Å	Internal contacts			IC_{50} (nM) m-calpain
					Good	Bad	Ugly	
A1	-7.4	-84.0	3	4.2	217	17	0	3710
A2	-7.4	-67.8	2	>5	274	26	0	
A3	-6.7	-77.3	3	3.8	254	8	0	
A4	-8.8	-77.3	3	4.9	258	5	0	
B1	-8.1	-53.5	3	3.9	287	16	1	280
B2	-7.0	-88.7	0	>5	287	21	0	
B3	-7.9	-69.0	3	4.4	261	10	1	
B4	-7.6	-81.6	0	>5	292	4	0	
C1	-6.0	-77.2	0	>5	300	18	0	
C2	-6.4	-78.6	3	4.1	307	31	6	

[a] Best pose chosen by criteria in order of importance; 1 - presence of three essential hydrogen bonds, 2 - warhead within 5Å of nucleophilic Cys115, 3 - low Glide score/emodel score, 4 - lowest number of internal ugly contacts.

[b] GlideScore[18] is a scoring function based on ChemScore[19] which is designed to estimate free energy of binding for protein–ligand complex. The function uses simple contact terms to estimate lipophilic and metal–ligand binding contributions, a simple explicit form for hydrogen bonds and a term which penalises flexibility.

[c] Emodel[18] is a model energy score that combines energy grid score, binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy.

Compound **A1** docks into the calpain InducedFit model with appropriate criteria met, namely the three essential H-bonds, a warhead-nucleophile distance of <5Å, and low Glide and Emodel scores (**Table 4.5**). It also showed no 'Ugly" internal contacts. In addition to the three essential H-bonds observed (left **Figure 4.11**) a fourth H-bond occurs between the aldehyde oxygen and Cys115. Lys347 has moved down towards the active site pocket to accommodate the ligand's capping group which has the 4-fluoro atom pointing towards the side chain of Trp224.

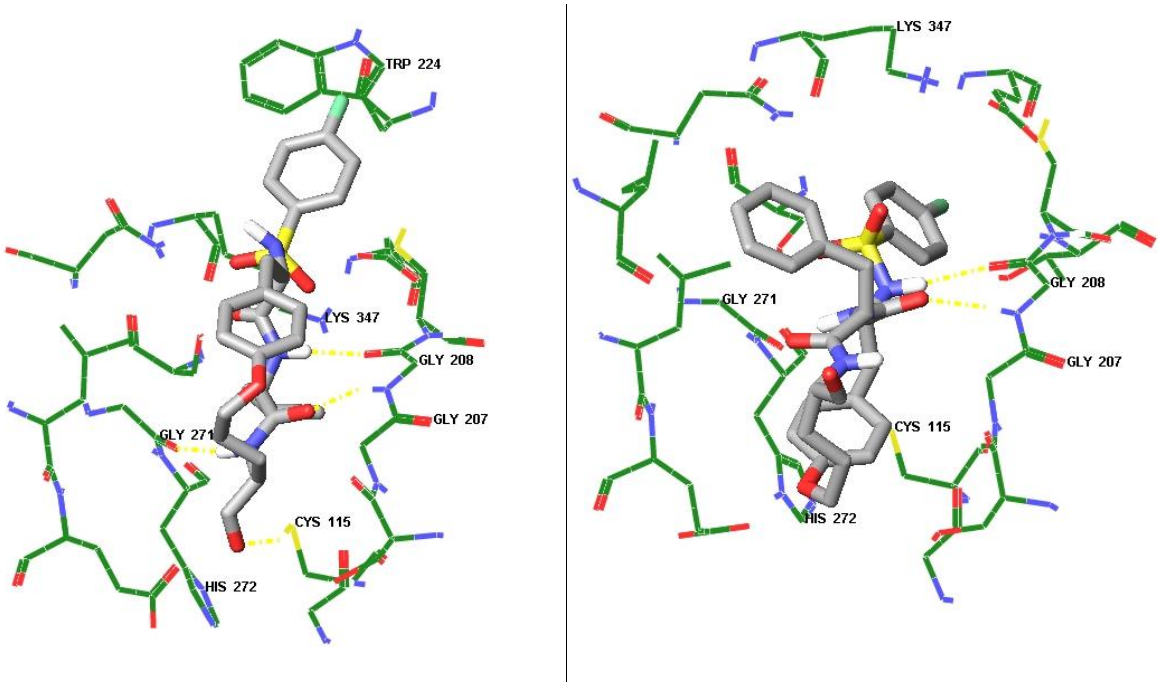


Figure 4.11; Best docked pose of compound **A1** (left) and compound **A2 (**right). All docked with the Induced Fit protocol into the InducedFit µ-calpain model

In parallel with the Glide docking study, compound **A2** does not bind in the required mode when the InducedFit protocol was employed (**Table 4.5** and right **Figure 4.11**). Again this underlines that the phenylalanine P_2 side chain hinders docking due to its bulk not being able to fit into the S_2 subsite.

Compounds **A3**, **A4**, and **B1** all show good InducedFit docking parameters (**Table 4.5**). They all have additional H-bonds present in their respective best docked pose. Compound **A3** has two extra H-bonds (**Figure 4.12**), one between the amide nitrogen at the capping group end and the side chain of Asn253, and another involving the aldehyde warhead oxygen of the ligand and the SH of Cys115.

Compound **A3** did not dock appropriately when docked using Glide, even from a *β*-strand starting conformation. The InducedFit protocol allows a small amount of movement in

the enzyme to accommodate the ligand so that it can bind in a manner akin to a 'good' inhibitor of calpain.

Compound **A4** has additional H-bonds, both from one of the oxygens of the sulfonyl capping group to the side chain of Asn253 and side chain of Ser251. It has an excellent Glide score of -8.8.

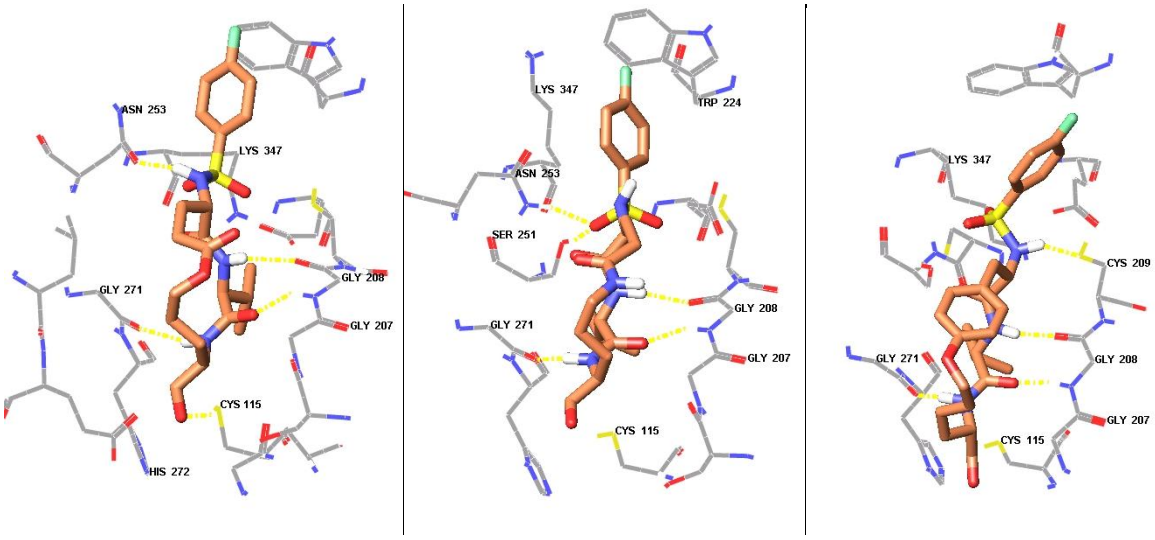


Figure 4.12: Best docked pose of compound **A3** (left), compound **A4** (middle), and compound **B1 (**right). All docked with the Induced Fit protocol into the InducedFit µ-calpain model

Compound **B1** has an additional H-bond from the NH group at the capping group end to the side chain sulphur of Cys209 (**Figure 4.12**). It has an excellent Glide score of -8.1 and a moderately good IC_{50} of 280 nM against m-calpain. It has only one 'ugly' internal contact. This compound failed to dock in the required manner with the Glide docking protocol. This could be due to the macrocycle ring of **B1** being larger than the analogous compound **A1** by one carbon bond. The InducedFit protocol allows enough movement in the enzyme model to allow **B1** to dock appropriately for inhibition.

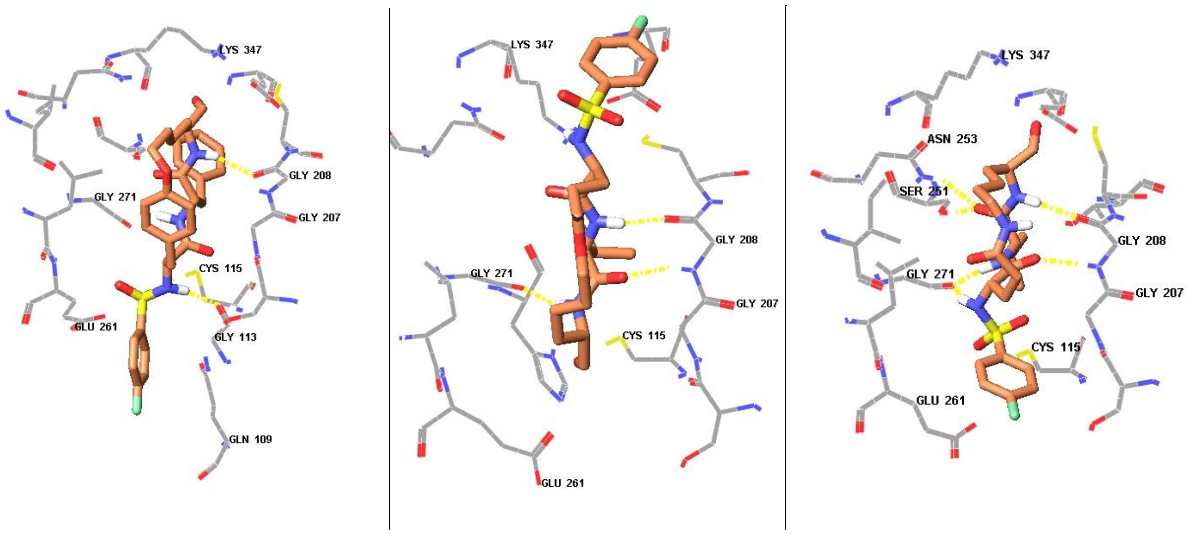


Figure 4.13: Best docked pose of compound **B2** (left), compound **B3** (middle), and compound **B4 (**right). All docked with the Induced Fit protocol into the InducedFit μ-calpain model

The InducedFit protocol failed to allow compound **B2** to dock into the model with the appropriate criteria (**Table 4.5** and left **Figure 4.13**). This was also the case with the Glide docking protocol and is considered to be due to the larger phenylalanine P_2 residue being unable to fit into the S_2 subsite of the enzyme. It was considered that the InducedFit protocol may permit the active site to flex enough to overcome the bulkier P_2 side chain and allow the ligand to dock with the correct criteria, but this was not the case.

Using the InducedFit protocol allowed compound **B3** to fit into the active site in an extended β-strand conformation with all required criteria bar one 'ugly' internal contact (**Table 4.5** and middle **Figure 4.13**). This was not the case when it was docked using the Glide protocol.

B4 was unable to dock appropriately using the Glide protocol or the InducedFit protocol (**Table 4.5** and right **Figure 4.13**).

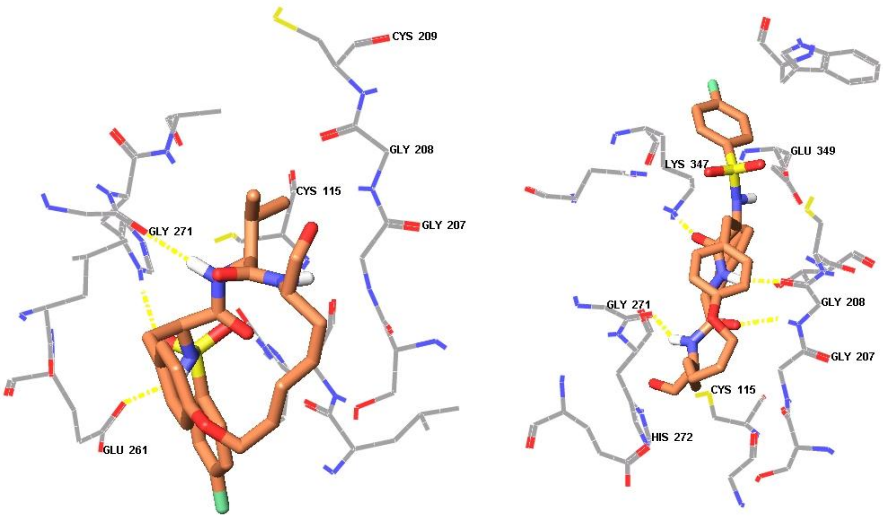


Figure 4.14; Best docked pose of compound **C1** (left) and compound **C2 (**right). All docked with the Induced Fit protocol into the InducedFit μ-calpain model

Compound **C1** docked with the Glide protocol in the required manner (**Figure 4.7**) but unexpectedly the InducedFit protocol did not allow the ligand to dock appropriately (**Table 4.5** and left **Figure 4.14**). Conversely, compound **C2** did not dock appropriately with the Glide protocol but did dock with the essential criteria when the InducedFit protocol was employed (**Table 4.5** and right **Figure 4.14**), however, the phenylalanine side chain of the ligand is twisted in order to fit into the S_2 subsite of the enzyme causing it to have a high number (six) 'ugly' internal contacts.

In summary, the Glide docking protocol takes up a lot of computer time and the InducedFit docking protocol takes up considerably more. To run the entire 288 compounds through these two docking protocols exceeded our computer time. For this reason we needed to test our methodology first.

It was for this reason that we tested our docking methodology on 10 macrocyclic compounds by first docking with Glide into the active site of the enzyme model. The data

collected from this initial docking study of the 10 selected compounds found that compounds **A1**, **A4**, and **C1** docked into the enzyme model in poses analogous to that found in compounds seen in co-crystallised X-ray structures found in the PDB[20-22], and in a way that adhered to our criteria for good inhibitors, namely having the three essential H-bonds, close warhead to nucleophile distance, and low Glide and Emodel scores. The other seven compounds could not dock in this way.

This initial Glide docking study revealed (by way of a visual inspection) that six of the seven compounds that did not dock appropriately were in poses that were not in the required *β*-strand conformation. An inspection of the lowest energy conformers used for docking revealed that these six compounds had their lowest energy conformer in a conformation that was not a *β*-strand. It was thought that using a non-*β*-strand conformer as a starting structure for Glide docking may result in these compounds not docking appropriately due to Glide not allowing for a rigorous search of conformational space for the potential inhibitor. The lowest energy conformer in a *β*-strand for each of these six compounds was then identified and Glide docking performed using these conformers. This allowed three more compounds (**B1**, **B2**, and **B3**) to dock with the essential criteria except for compound **B2** missing an essential H-bond resulting from the large P_2 side chain being unable to fit into the deep S_2 subsite and resulting in the ligand being positioned higher and out of the active site.

Overall, six out of the ten compounds (**A1**, **A4**, **B1**, **B2**, **B3** and **C1**) could dock in the required manner when the lowest energy conformer in a *β*-strand for each compound was used as the starting conformer in the Glide docking study. However, some of these compounds were found to have a small number of ‘ugly’ internal contacts which were thought to be a consequence of Glide treating the enzyme model as a rigid structure (this is

not the case in nature) which forces the large macrocycle compounds to dock in a slightly 'forced' manner. It was thought that if the enzyme model was allowed to 'relax' a little that this would also allow the docked compounds to 'relax' and this would see the elimination of some or all of the 'ugly' internal contacts. It was for this reason that Schrodingers InducedFit protocol was employed. It was also hoped that the InducedFit protocol would allow compounds such as **A2** with a large phenylalanine side chain and compounds with larger macrocycle rings to position into the active site pocket appropriately.

The ten compounds used in the initial Glide docking study were then docked using the InducedFit docking protocol into the calpain enzyme model. Of the compounds **A1**, **A2**, **A3**, and **A4** only **A2** did not display the essential criteria. As discussed earlier it was considered that the InducedFit 'movement' of the enzyme model would allow the enzyme to be more flexible and would help overcome compound **A2**'s larger P_2 side chain and allow the compound to fit into enzyme model but this proved not to be the case.

Of the compounds **B1**, **B2**, **B3**, and **B4** only **B2** and **B4** did not dock appropriately. **B2** is similar to **A2** as it has a phenylalanine P_2 side chain and it is likely it suffers from that same problem as **A2**. **B4** is analogous to **A4** but has a larger macrocycle ring by one bond and it is probable that this larger ring size causes the compound not to dock in the required manner due to its size.

While **C1** docked appropriately in the initial Glide docking it did not dock in the same way when the InducedFit protocol was used. Although **C2** was able to dock with the essential criteria when InducedFit was used, a high number of 'ugly' internal contacts were evident

due to the phenylalanine side chain being twisted in order to fit into the deep S_2 subsite of the enzyme.

The results of the InducedFit docking studies on compounds **A1**-**A4**, **B1**-**B4**, **C1** and **C2** (**Table 4.5**) indicated that compounds **A1**, **A3**, **A4**, **B1**, and **B3** are potentially good inhibitors of calpain. Two of these compounds (**A1** and **B1**) were prepared based on the modelling results and availability of the required chemicals for synthesis. They were tested in house and were shown to inhibit calpain *in vitro* with IC_{50}'s of 3710 and 280 nM, respectively. A valine side chain at the P_2 position appears to be a better candidate than a phenylalanine at the same position as the latter seems to hinder docking of macrocycles **A2**, **B2**, and **C2**.

As discussed above, the Glide and InducedFit docking protocols take considerable computer time on the basis of the above study of 10 macrocyclic compounds. It was therefore decided that the 288 compound library would first be examined by analysis of the conformational search outputs to establish the compounds with a marked propensity for *β*-strand conformations. From this study the best were selected for subsequent docking with the InducedFit Protocol.

4.3.3 Determination of Boltzmann weighted percentage of *β*-strand conformers within the macrocycle library of low energy ensembles

Earlier observations revealed that not all of the lowest energy conformers for each of the 288 compounds were in a *β*-strand. The occurrence of a *γ*-turn was unexpected and problematic because on further examination it was noticed that many of the ensembles contained mixtures of *γ*-turns and *β*-strand. Docking of *γ*-turn conformers causes a problem because within the small binding pocket the rotational freedom is more restricted. The Glide

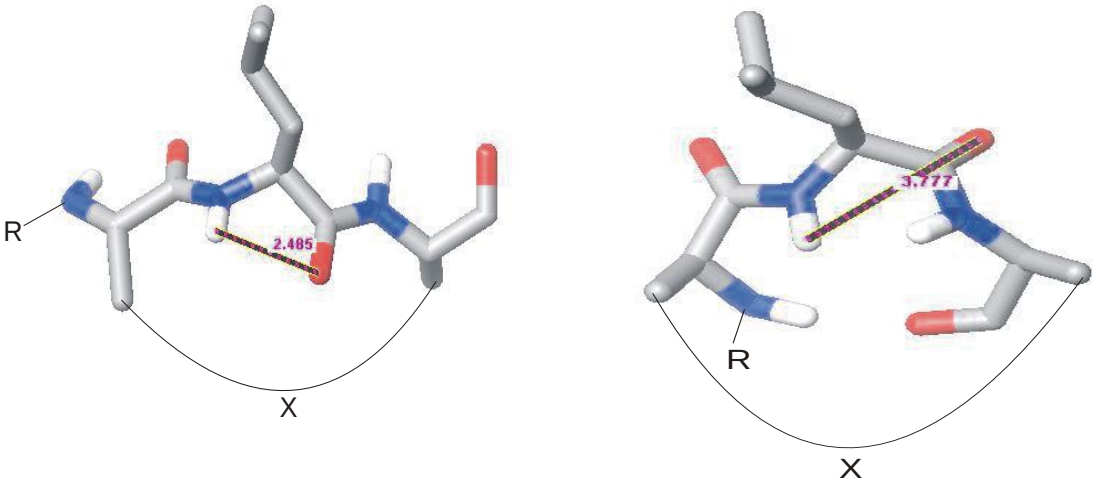
docking program does not allow the peptide backbone to 'flip' from a γ-turn to a β-strand (and vice versa) within the binding pocket. Therefore, docking a macrocycle in a γ-turn conformation gives docking poses with the macrocycle only in a γ-turn conformation.

It was necessary to conduct an analysis of the 288 macrocycles in the virtual library to identify appropriate conformers to use to initiate docking. By determining the Boltzmann weighted percentage of β-strand for each macrocycles low energy ensemble we could eliminate compounds that do not prefer a β-strand conformation while identifying compounds that do favour a β-strand. This was thought to be an important step because a compound that could not form a β-strand would in theory not be able to bind the active site of calpain and consequently effect inhibition of the enzyme. Compounds that could form a β-strand but do not prefer a β-strand conformation would in theory be less effective inhibitors than compounds that did prefer such a bioactive conformation.

Firstly, a visual inspection of the ensembles of conformers of each macrocycle was conducted along with an analysis of some *in silico* interatomic measurements which were used to define parameters (albeit arbitrary) to allow a turn, a strand, and a 'twist' conformer to be identified.

For example, the interatomic distance between the P_2 backbone oxygen and the P_2 backbone NH hydrogen in the left picture of **Figure 4.15** measures 2.485 Å and is indicative of a β-strand and the analogous interatomic distance of 3.777 Å in the right picture is indicative of a γ-turn. The parameter I used is as follows; If the interatomic distance between the P_2 backbone oxygen and the P_2 backbone NH hydrogen is < 3.1 Å the macrocycle was

identified as a β-strand, when > 3.6 Å a γ-turn, and between 3.1 and 3.6 Å a 'twist' conformer. These interatomic distances are easily tabulated for all ensemble conformers.



A. **B**

Figure 4.15. A Macrocycle with typical β-strand conformation showing an interatomic distance between the carbonyl O atom and the amino H atom of the central leucine residue of 2.485 Å. **B.** Macrocycle with typical γ-turn conformation showing an interatomic distance measurement of 3.777 Å. R can be any capping group such as BOC. X can be any two side-chains able to be joined, for example, by ring closing metathesis.

I attempted to calculate the Boltzmann weighted percentage of strand, turn, and 'twist' conformers for each low energy ensemble of macrocycle using this method.

In order to calculate this data, a spreadsheet was established that would calculate the Boltzmann weighted percentages when the potential energy and interatomic parameters for each conformer was entered. An example of this spreadsheet is shown below in **Table 4.6**.

Table 4.6. A section of the spreadsheet used to calculate the Boltzmann weighted percentages of conformers exhibiting turn, strand, and 'twist'.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Energy	E/kT	e-E/kT	Ne-E/kT	Boltzmann	%	E*%	Distance 1	Bolt.Ave.E	Strand	Turn	Twist %	Strand %	Turn %
-316.54	126.9	###	###	2.23	24.76	-7837.15	2.28	-314.9	24.76	0	0	100	0
-316.06	126.71	###	###	1.84	20.4	-6448.01	2.2		20.4	0			
-314.42	126.05	###	###	0.95	10.6	-3332.98	2.19		10.6	0			
-313.9	125.85	###	###	0.77	8.61	-2702.56	2.13		8.61	0			
-313.9	125.84	###	###	0.77	8.59	-2697.93	2.18		8.59	0			
-313.87	125.83	###	###	0.76	8.49	-2665.43	2.55		8.49	0			
-313.69	125.76	###	###	0.71	7.92	-2483.7	2.18		7.92	0			
-312.73	125.38	###	###	0.48	5.38	-1684.01	2.17		5.38	0			
-312.66	125.35	###	###	0.47	5.24	-1638.39	2.12		5.24	0			

Column title meanings:

A	Potential energy of each conformer calculated by the OPLS2005 force field.
B, C, D, E	Each column calculates the Boltzmann weighting of each conformer from the expression $$P_\alpha = \frac{\exp[-(E_\alpha / k_B T]}{\sum_{\alpha^-}^{N_A} \exp[-(E_\alpha / k_B T]}$$
B	E/kT where E = negative potential energy (the negative of column A), k = Boltzmann constant (8.32441 $J \cdot K^{-1} \cdot mol^{-1}$), T = temperature (300K). Formula =(-A2*1000)/(8.31451*300)
C	e-E/kT, e = natural log. Formula =IF(B2=0,0,EXP(B2))
D	Ne-E/kT, N = number of conformers (entries in column A). Formula =COUNT($A$2:$A$300)*(C2)
E	Boltzmann. Final calculation performed to give the Boltzmann weighting of each conformer (entries in column A). Formula =(D2)/SUM($C$2:$C$300)
F	%. Calculates the Boltzmann weighted % of each conformer. Formula =(E2)/SUM($E$2:$E$300)*10
G	E*%. Potential energy (column A) x Boltzmann weighted % (column F). Formula =(A2*F2)
H	Interatomic distances between the O atom and H atom indicated in Figure 2.
I	Bolt.Ave.E. Calculates the Boltzmann average potential energy within 12 kJ of the lowest energy conformer. Formula =SUM(G2:G300)/100
J	Calculates the Boltzmann weighting of conformers exhibiting a β-strand. Formula =IF(H3>3.1;0;F3).
K	Calculates the Boltzmann weighting of conformers exhibiting a γ-turn. Formula =IF(H3<3.6;0;F3).
L	Calculates the Boltzmann weighted percentage of conformers exhibiting a 'twist' Formula =100-(SUM($J$2:$J$1046))-(SUM($K$2:$K$1046)).
M	Calculates the Boltzmann weighted percentage of conformers exhibiting a β-strand. Formula =SUM($J$2:$J$1046).
N	Calculates the Boltzmann weighted percentage of conformers exhibiting a γ-turn. Formula =SUM($K$2:$K$1046)

The method worked well but was time consuming as the potential energies and the interatomic distances had to be cut and pasted into a spreadsheet for each macrocycle. A script was subsequently written with the help of Dr. Quentin MacDonald to calculate the Boltzmann weighted percentage of *β*-strand for each macrocycle. The script was written to identify *β*-strands based on the *ψ* (Psi) and *φ* (Phi) angles of the P_2 amino acid of each compound. Ramachandran plots of *ψ*,*φ* angles in *β*-strand or *β*-sheet regions of protein X-ray crystal structures from the PDB show that typical *ψ* angles to be between 90º and 160º and that of *φ* angles to be between -90º and -160º.[23] These parameters were used in the program to

determine the Boltzmann weighted percentage of *β*-strand for each macrocycle. The program script was run using the output files from the conformational searches performed on the macrocycles and the resulting data tabulated (**Table 4.7)**. The table shows good candidates (green cells), average candidates (uncoloured cells), and poor candidates (red cells) depending on the Boltzmann weighted percentage of each macrocycle in a *β*-strand. It also has the ring size of each macrocycle at the bottom of each cell.

Table 4.7. Each cell contains data relating to the equivalent 4-fluorobenzyl sulphonamide protected aldehydes **A1**, **A2**, **A3**...**P16**, **P17**, **P18**. The top of each cell shows the Boltzmann weighted percentage of each macrocycle in a β-strand, green cells shows macrocycles with >70% β-strand, uncoloured cells show between 30% and 70% β-strand, and red cells show <30% β-strand. The bottom of each cell shows the macrocycle ring size by number of bonds.
[a] Compounds **A8** and **B7** are equivalent structures.
[b] Compounds **A9**, **B8**, and **C7** are equivalent structures.
[c] Compounds **B9** and **C8** are equivalent structures.

								Acid										
Ester	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
A	100	100	16	50	0	2	0	0	0	1	29	67	61	67	0	0	43	98
	16	16	14	14	13	13	9	10[a]	11[b]	12	13	14	15	16	12	12	14	16
B	57	27	8	6	28	2	0	0	2	24	21	29	88	63	1	7	74	39
	17	17	15	15	14	14	10[a]	11[b]	12[c]	13	14	15	16	17	13	13	15	17
C	53	16	78	0	24	2	0	2	12	33	0	72	80	75	0	0	81	37
	18	18	16	16	15	15	11[b]	12[c]	13	14	15	16	17	18	14	14	16	18
D	28	15	0	41	3	1	68	29	28	20	21	7	7	5	34	17	1	2
	21	21	19	19	18	18	14	15	16	17	18	19	20	21	17	17	19	21
E	47	3	0	0	1	11	0	60	26	0	0	1	15	0	24	0	0	18
	21	21	19	19	18	18	14	15	16	17	18	19	20	21	17	17	19	21
F	63	23	0	0	15	0	27	37	43	10	27	0	48	18	29	5	36	13
	20	20	18	18	17	17	13	14	15	16	17	18	19	20	16	16	18	20
G	35	5	2	0	1	6	1	11	49	0	0	3	6	58	20	5	8	2
	20	20	18	18	17	17	13	14	15	16	17	18	19	20	16	16	18	20
H	95	47	7	0	68	32	0	16	65	0	0	38	30	4	83	18	25	1
	19	19	17	17	16	16	12	13	14	15	16	17	18	19	15	15	17	19
I	32	75	41	0	12	4	34	30	98	0	80	63	26	8	26	0	24	13
	20	20	18	18	17	17	13	14	15	16	17	18	19	20	16	16	18	20
J	0	17	11	31	29	1	2	13	42	0	6	0	19	2	0	0	40	0
	21	21	19	19	18	18	14	15	16	17	18	19	20	21	17	17	19	21
K	10	0	43	0	22	0	85	89	71	0	0	2	55	1	0	0	0	13
	22	22	20	20	19	19	15	16	17	18	19	20	21	22	18	18	20	22
L	52	5	0	0	0	3	1	2	9	0	0	0	53	0	0	0	0	0
	23	23	21	21	20	20	16	17	18	19	20	21	22	23	19	19	21	23

M	4	0	0	0	6	0	0	0	1	0	0	25	68	0	3	0	38	0
	19	19	17	17	16	16	12	13	14	15	16	17	18	19	15	15	17	19
N	6	0	0	20	17	21	0	0	0	0	0	1	0	50	0	0	0	0
	19	19	17	17	16	16	12	13	14	15	16	17	18	19	15	15	17	19
O	8	6	88	21	28	0	0	0	60	0	0	0	11	0	4	0	0	0
	21	21	19	19	18	18	14	15	16	17	18	19	20	21	17	17	19	21
P	61	0	0	0	2	0	92	94	0	8	0	3	0	3	1	1	0	4
	23	23	21	21	20	20	16	17	18	19	20	21	22	23	19	19	21	23

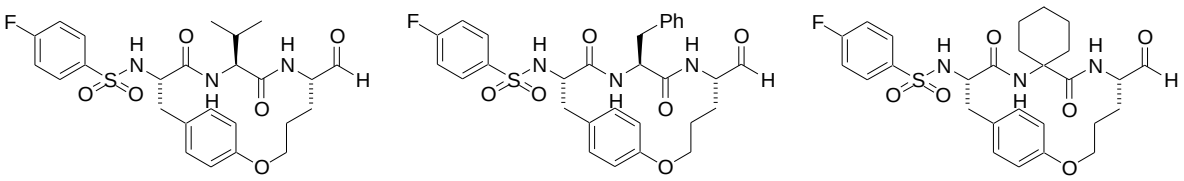
Twenty compounds that were found to have a more than 70% Boltzmann weighted preference for a *β*-strand conformation (green cells) are shown in **Table 4.7**. Seven of these have a 90% Boltzmann weighted preference for a *β*-strand conformation. These seven were chosen to be docked into the enzyme model using the InducedFit protocol.

For docking of these compounds the InducedFit Protocol was chosen as it was hoped that it might allow some of the larger compounds to fit into the active site which had been shown in the previous study. It had already been shown that compound **A2** could not be docked into the rigid enzyme model using Glide alone and that this was likely due to the bulkier side chain compared to the similar compound **A1** as discussed earlier.

The 20 compounds showing a 70% preference for a *β*-strand all have a ring size of between 15 and 19 bonds which is consistent with Fairlies[6] assessment that ca 17 membered macrocycles prefer a *β*-strand conformation. Many of the compounds with smaller and larger ring sizes do not form a *β*-strand at all.

4.3.4 InducedFit docking of seven compounds with a >90% preference for a *β*-strand conformation

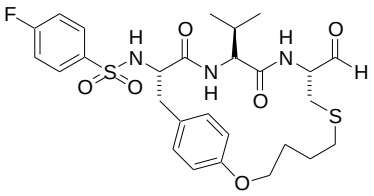
Compounds **A1**, **A2**, **A18**, **H1**, **I9**, **P7**, and **P8** (**Figure 4.16**) all have a >90% preference for a *β*-strand conformation and were docked using the lowest energy conformer in a *β*-strand and Schrodinger's InducedFit Protocol (parameters used are shown in **Protocol 5** in the **Appendix**). Compounds **A1** and **A2** were actually docked with InducedFit earlier as part of the docking study in **Section 4.1.5** so will only be discussed briefly here.



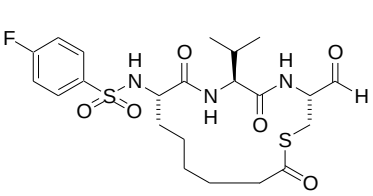
A1

A2

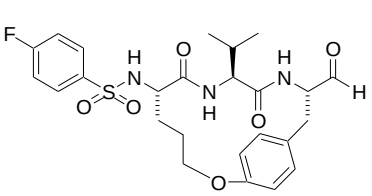
A18



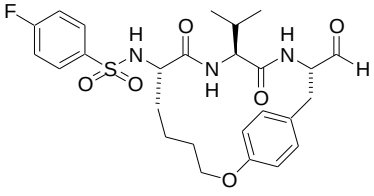
H1



I9



P7



P8

Figure 4.16: Chemdraw pictures of compounds **A1**, **A2**, **A18**, **H1**, **I9**, **P7**, and **P8**

Table 4.8: Docking data for best pose (out of possible 20) and inhibitory concentrations (IC_{50}) for compounds **A1**, **A2**, **A18**, **H1**, **I9**, **P7**, and **P8**

Compound	Glide Score [b]	emodel Score [c]	Essential H bonds	Warhead Distance Å	Internal contacts			IC_{50} (nM) m-calpain
					Good	Bad	Ugly	
A1	-7.4	-84.0	3	4.2	217	17	0	3710
A2	-7.4	-67.8	2	>5	274	26	0	
A18	-6.4	-69.6	2	3.5	378	31	13	
H1	-7.6	-87.3	3	3.5	295	15	0	2400
I9	-6.9	-68.5	3	3.5	271	7	1	
P7	-5.1	-71.9	3	3.8	285	11	0	
P8	-8.0	-70.2	3	4.0	291	12	0	

Compounds **A1** and **A2** were docked using the InducedFit Protocol. Compound **A1** docks into the calpain InducedFit model with all the essential criteria in place. Compound **A2** could not bind in the required mode when the InducedFit protocol was employed. For further discussion on docking of compounds **A1** and **A2** see **Section 4.1.5** above (**Table 4.5** and **Figure 4.11**).

Compounds **H1** and **P7** gave docked poses that had all the essential requirements for good calpain inhibition (**Table 4.8** and **Figure 4.17**). In addition to having the three required H-bonds they both had two additional H-bonds. Both compounds had one of their extra H-bonds formed between a sulfonyl oxygen from the capping group and the side chain of Asn253. The compounds also have an additional H-bond formed between the warhead aldehyde oxygen and the side chain of His272. These extra H-bonds would increase the binding affinity in a real biological system.

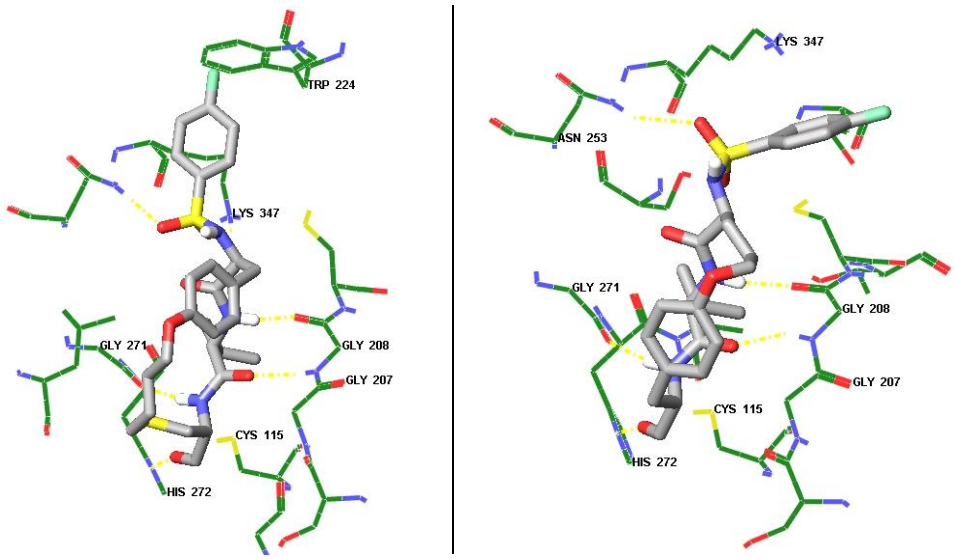


Figure 4.17; Best docked pose of compound **H1** (left) and compound **P7 (**right). All docked with the Induced Fit protocol into the InducedFit μ-calpain model

The additional data for compounds **H1** and **P7** is also encouraging as both have a close distance between the warhead and nucleophile (3.5 and 3.8, respectively), and good Glide and Emodel scores, and their best poses display no 'ugly' internal contacts.

Figure 4.18 shows the best docked poses of **P8** (left), **I9** (middle), and **A18 (**right). **P8** and **I9** have the three essential H-bonds and all other necessary criteria for good calpain inhibition (**Table 4.8**). Compound **P8** also has an additional H-bond formed between a sulfonyl oxygen from the capping group and the side chain of Asn253.

Compound **A18** seems to be suffering from the same problem as **A2**, **B2**, and **C2** in the previous docking study because of its bulkier P_2 side chain, in this case a cyclohexanyl group. This group only partially fits into the S_2 subsite of the enzyme models active site but the β-strand backbone of the ligand is sitting higher in the pocket and consequently one of the essential three H-bonds cannot form.

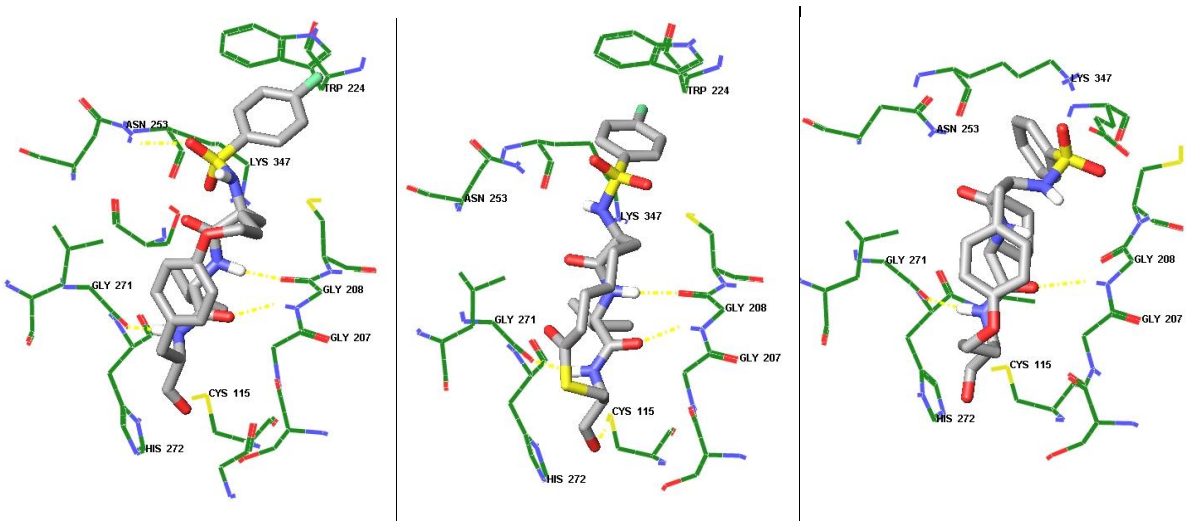


Figure 4.18; Best docked pose of compound **P8** (left), compound **I9** (middle), and compound **A18 (**right). All docked with the Induced Fit protocol into the InducedFit μ-calpain model

In summary, the compounds **A1**, **A2**, **A18**, **H1**, **I9**, **P7**, and **P8** from the 288 compound library were selected for InducedFit docking studies because their low energy conformational ensembles showed each of these compounds had a high propensity for having β-strand conformations. All had a greater than 90% Boltzmann weighted preference for a β-strand (**Table 4.7**).

As previously discussed compound **A1** had all the docking criteria indicating it to be an active inhibitor of calpain, while compound **A2** did not dock in the required way because of it having a larger P_2 side chain. Compound **A18** suffered from a similar problem to **A2** with having a bulky P_2 side chain that prevented it from docking with all three essential H-bonds. Like compound **A1**, compounds **H1**, **I9**, **P7**, and **P8** had good InducedFit docking results and showed promise as possible inhibitors of calpain.

As a result of these docking studies, compounds **A1**, **B1**, and **H1** were chosen as targets for synthesis by members of the cataract research team. These compounds were subsequently

tested *in vitro* and were shown to inhibit ovine m-calpain at 3710, 280, and 2400 nM, respectively.

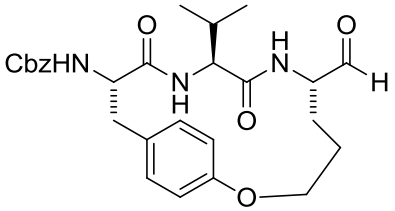
4.3.5 InducedFit docking: Analogues of macrocyclic compounds A1, B1, C1, and H1

To further our study on the macrocycle compounds, the macrocycle core of compounds **A1**, **B1**, **C1** and **H1** were used as templates to add a different capping group, a different P_2 side chain, and a different warhead. The templates of compounds **A1**, **B1**, and **H1** were chosen because they had been shown to be inhibitors of calpain with varying degrees of potency in the previous study. The template of **C1** was chosen as it was an analogous molecule to **A1** and **B1** but had a larger macrocycle ring and docking data of this compound had previously been inconclusive as it had seemingly docked well with Glide but had failed to dock appropriately when InducedFit was used (**Sections 4.1.4** and **4.1.5**, respectively). For the following study a Cbz group was chosen as an alternative capping group for a series of compounds with the above mentioned templates.

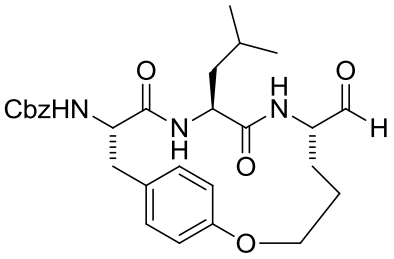
The previous sections showed that phenylalanine at the P_2 position to be too large to fit into the deep hydrophobic pocket of the S_2 subsite while a smaller valine at this position proved to be a better fit. As a compromise, leucine, which is larger than valine but smaller than phenylalanine and still a hydrophobic amino acid, was investigated at the P_2 position.

To overcome the problems previously discussed with respect to the aldehyde warhead being too reactive with other cellular proteins[10] to be considered as a good drug candidate we decided to replace the aldehyde with an alcohol. An alcohol was thought to be a good

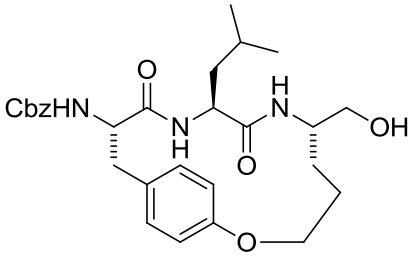
candidate because, although not a warhead as such, it would mimic the tetrahedral intermediate formed when the aldehyde reacts with the active site cysteine of the calpain enzyme, as is seen in PDB X-ray structures like 2G8E,[20] and although these compounds would not form a reversible covalent bond with the active site cysteine the trade off with them not reacting with other proteins in a biological system was thought to be worth examining. These compounds can be seen in **Figure 4.19**



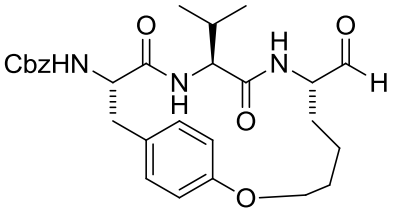
4.9



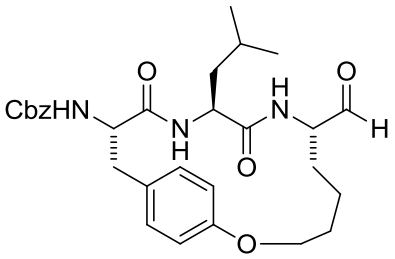
4.10



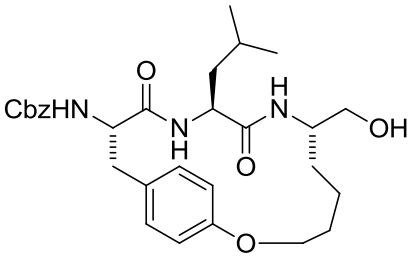
4.11



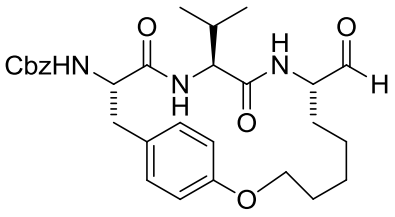
4.12



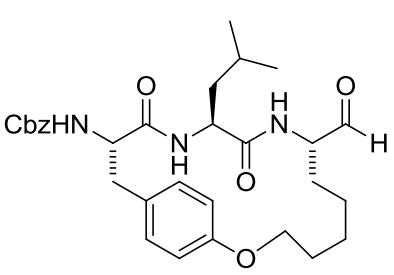
4.13



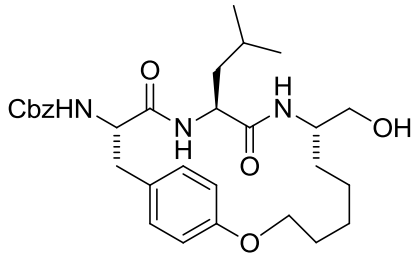
4.14



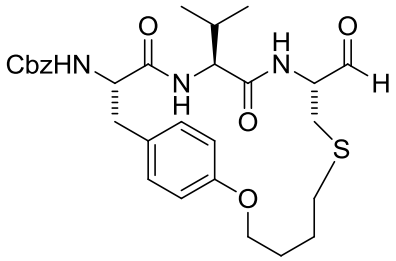
4.15



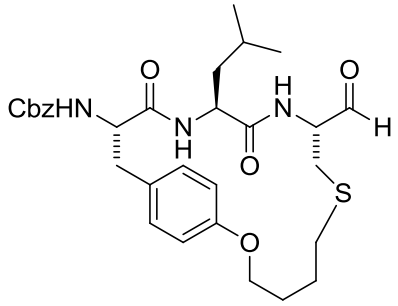
4.16



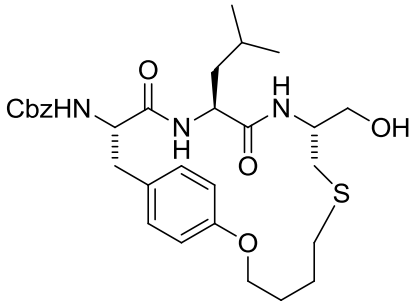
4.17



4.18



4.19



4.20

Figure 4.19: Structures of compounds **4.9-4.20**

Table 4.9: Docking data for best pose[a] (out of possible 20) and inhibitory concentrations (IC_{50}) for compounds **4.9-4.20**

Compound	Glide Score [b]	emodel Score [c]	Essential H bonds	Warhead Distance Å	Internal contacts			IC_{50} (nM) m-calpain
					Good	Bad	Ugly	
4.9	-8.3	-71.3	3	3.4	277	17	1	
4.10	-8.4	-86.6	3	3.7	307	22	2	850
4.11	-9.7	-83.7	3	NA	311	26	5	31000
4.12	-8.0	-78.1	3	3.9	280	18	1	85
4.13	-4.5	-67.3	3	4.0	309	20	1	30
4.14	-9.0	-77.5	3	NA	340	24	2	700
4.15	-7.8	-84.9	3	4.1	303	18	0	
4.16	-10.0	-69.1	3	4.0	325	14	3	180
4.17	-10.1	-82.8	3	NA	313	21	1	1100
4.18	-9.2	-87.5	3	4.2	285	14	1	295
4.19	-9.7	-66.4	3	4.8	314	21	2	1010
4.20	-9.4	-79.3	0	NA	347	26	8	28000

[a] best pose chosen by criteria in order of importance; 1 - presence of three essential hydrogen bonds, 2 - warhead within 5Å of nucleophilic Cys115, 3 - low Glide score/emodel score, 4 - lowest number of internal ugly contacts.

[b] GlideScore[18] is a scoring function based on ChemScore[19] which is designed to estimate free energy of binding for protein–ligand complex. The function uses simple contact terms to estimate lipophilic and metal–ligand binding contributions, a simple explicit form for hydrogen bonds and a term which penalises flexibility.

[c] Emodel[18] is a model energy score that combines energy grid score, binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy.

Compounds **4.9**, **4.10**, and **4.11** docked into the enzyme model with all the essential criteria (**Table 4.9** and **Figure 4.20**). These compounds had *β*-strand percentages of 99%, 100%, and 99%, respectively, showing that they strongly preferred this conformation (**Table 4.10**). In addition to the three essential H-bonds, compound **4.10** had an extra H-bond between the amide oxygen at the capping group end to the side chain of Asn253. The IC_{50} for **4.10** was recorded at a moderately good 850 nM.

Compounds **4.9** and **4.10** had only one and two 'ugly' internal contacts, respectively. However, compound **4.11** displayed five 'ugly' internal contacts which is a possible indication that it was not a good fit into the enzyme and may be why the IC_{50} for this compound turned out to be particularly high (31000 nM). However, a lower IC_{50} is expected for the compounds with an alcohol 'warhead' as they do not form a covalent bond.

Table 4.10: The Boltzmann weight percentage of conformers within a 12kJ window displaying a *β*-strand for compounds **1-12** and the macrocycle ring size for each compound.

	Compounds											
	4.9	**4.10**	**4.11**	**4.12**	**4.13**	**4.14**	**4.15**	**4.16**	**4.17**	**4.18**	**4.19**	**4.20**
% *β*-strand	99	100	99	65	8	75	89	85	53	35	39	0
Macrocycle ring size	16	16	16	17	17	17	18	18	18	19	19	19

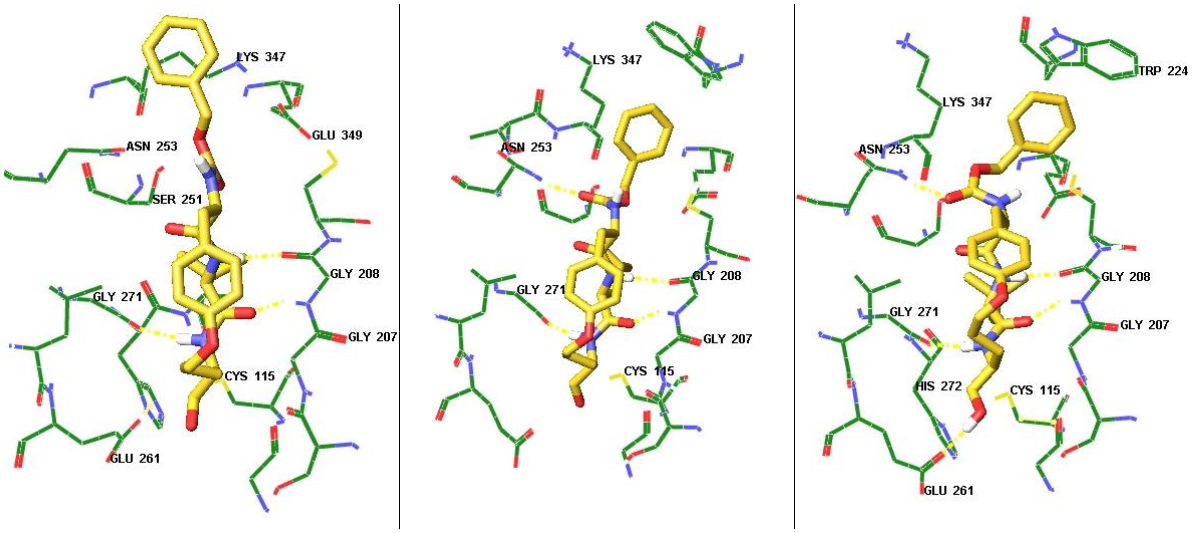


Figure 4.20: Best docked poses of compounds **4.9**, **4.10**, and **4.11** using the InducedFit protocol docked into the calpain model.

Compounds **4.12**, **4.13**, and **4.14** also demonstrated a preference to bind in the appropriate way (**Table 4.9** and **Figure 4.21**). Extra H-bonds were seen in the best docked pose of compound **4.12**, one between the aldehyde warhead oxygen and the side chain of Cys115, and the other from the backbone oxygen of the P_3 residue to the side chain of Asn253. When tested *in vitro* compound **4.12** had an excellent IC_{50} of 85 nM, one of the most potent in the series. It also has a ring size of 17 which fits with Fairlie's[6] assessment of having a ca 17 membered ring to favour a *β*-strand.

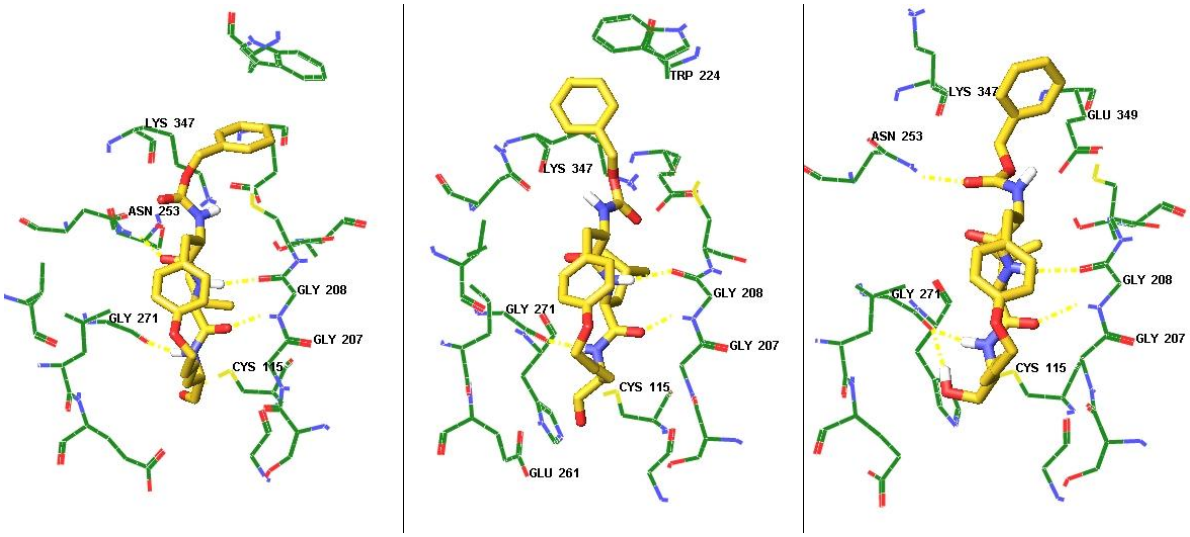


Figure 4.21: Best docked poses of compounds **4.12**, **4.13**, and **4.14** using the InducedFit protocol docked into the calpain model.

Compound **4.13** was shown to be the most potent compound *in vitro* with an excellent IC_{50} of 30 nM against ovine m-calpain (**Table 4.9**). The larger leucine side chain (compared to valine) at the P_2 position fits into the deep hydrophobic pocket of the S_2 subsite without problem (**Figure 4.21**). Changing from a valine (compound **4.12**) to a leucine (compound **4.13**) has had a positive effect on the IC_{50}. It has the optimal macrocyclic ring size of 17.

Compound **4.14** has an IC_{50} of 700 nM which is particularly good considering it has an alcohol in place of the usual aldehyde warhead. The alcohol forms a H-bond with the side chain of Gly271 (**Figure 4.21** left) and is unexpectedly pointing away from the oxyanion hole (usually formed by two stabilising H-bonds, one between the oxyanion of a tetrahedral intermediate and the backbone NH of Cys115 and the other between the oxyanion and the side chain of Gln109).[20]

Compounds **4.15**, **4.16**, and **4.17** all bind to the active site pocket with the required criteria. All have two extra H-bonds in addition to the three essential H-bonds required for good binding (**Figure 4.22** left). Compound **4.16** has three 'ugly' internal contacts but when synthesised and tested against m-calpain it has a good IC_{50} of 180 nM suggesting that the three 'ugly' contacts could be accommodated by the enzyme.

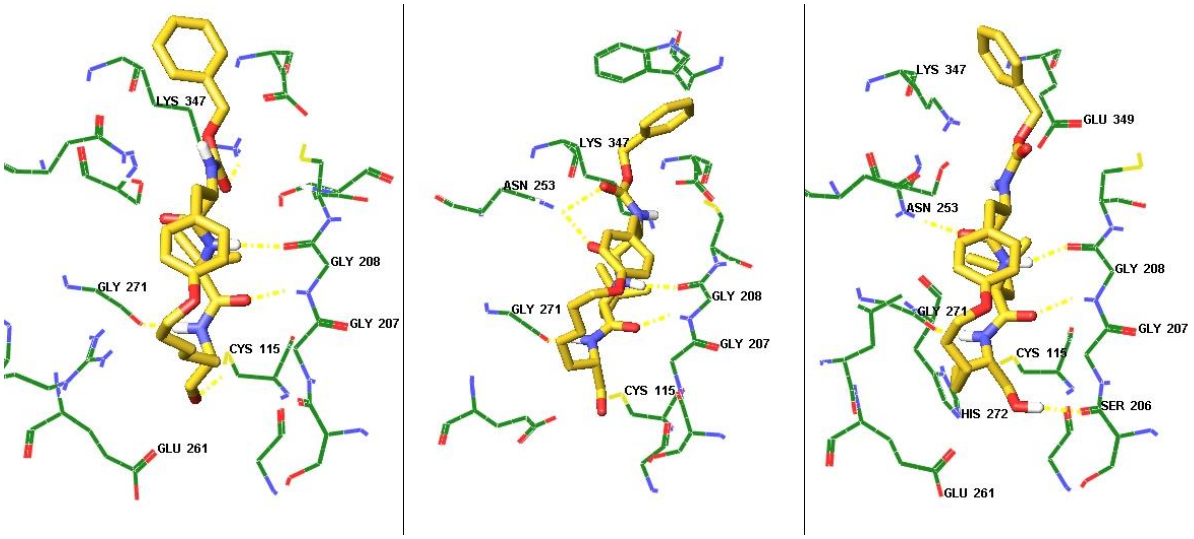


Figure 4.22: Best docked poses of compounds **4.15**, **4.16**, and **4.17** using the InducedFit protocol docked into the calpain model.

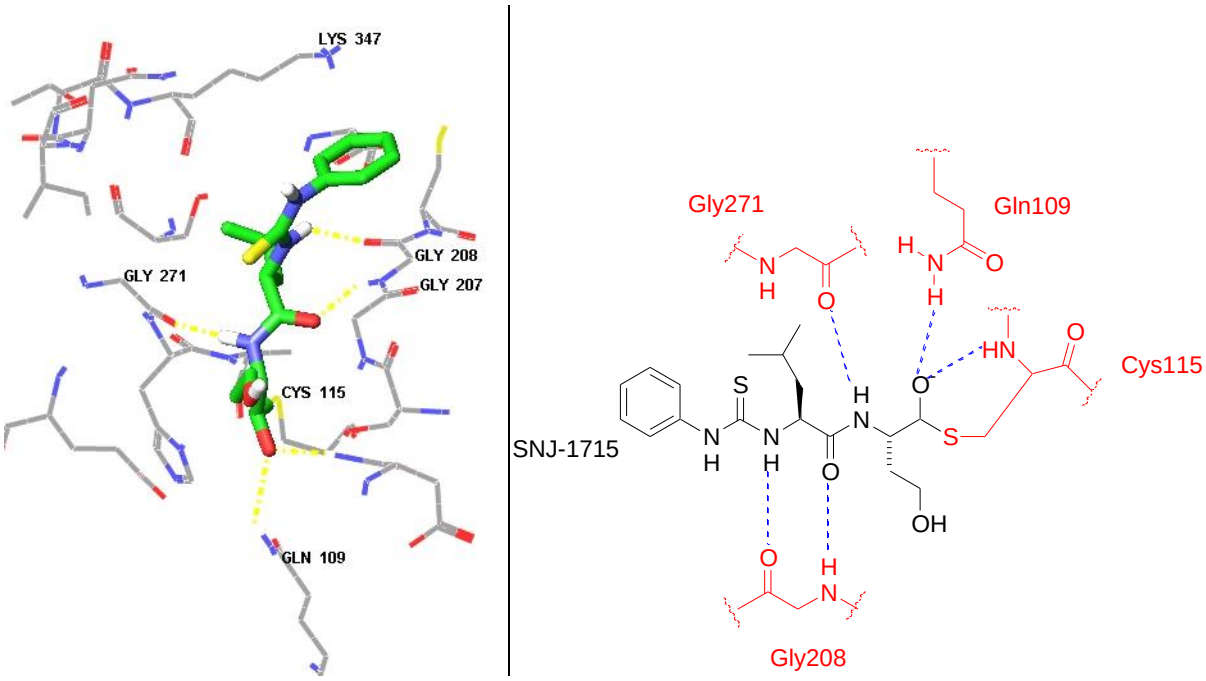


Figure 4.23: Left; Picture of X-ray crystal of compound SNJ-1715 bound to mini μ-calpain (PDB code 2G8E)[20] showing the three essential H-bonds to Gly208 and Gly271 and the two stabilising H-bonds from the oxyanion to the side chain of Gln109 and the backbone NH of Cys115. Right; Chemdraw depiction of SNJ-1715 bound to the active site of μ-calpain (PDB code 2G8E).[20] Enzyme (red), inhibitor (black), and H-bonds (blue).

Compound **4.17** has an alcohol at the 'warhead' end and its best docked pose shows that it is positioned closely to the oxyanion hole but forming a H-bond to the side chain of Gly113 instead of Cys115 and Gln109. It cannot form the usual H-bonds to Cys115 and Gln109 as it is positioned higher in the pocket than is usually seen when an aldehyde or similar type warhead covalently binds to the active site cysteine and the oxyanion of the tetrahedral intermediate forms these two H-bonds as seen in the PDB X-ray structure 2G8E (**Figure 4.23** right).[20]

Compound **4.18** fitted into the active site with all the required criteria (**Table 4.9** and **Figure 4.24**). It had two extra H-bonds, one from the carbonyl oxygen at the capping group end to Asn253 and the other from the aldehyde warhead to the side chain of Cys115. It was

synthesised and tested and gave an IC_{50} of 195 nM. **4.18** has a valine at the P2 position while the analogous compound **4.19** has a leucine at this position.

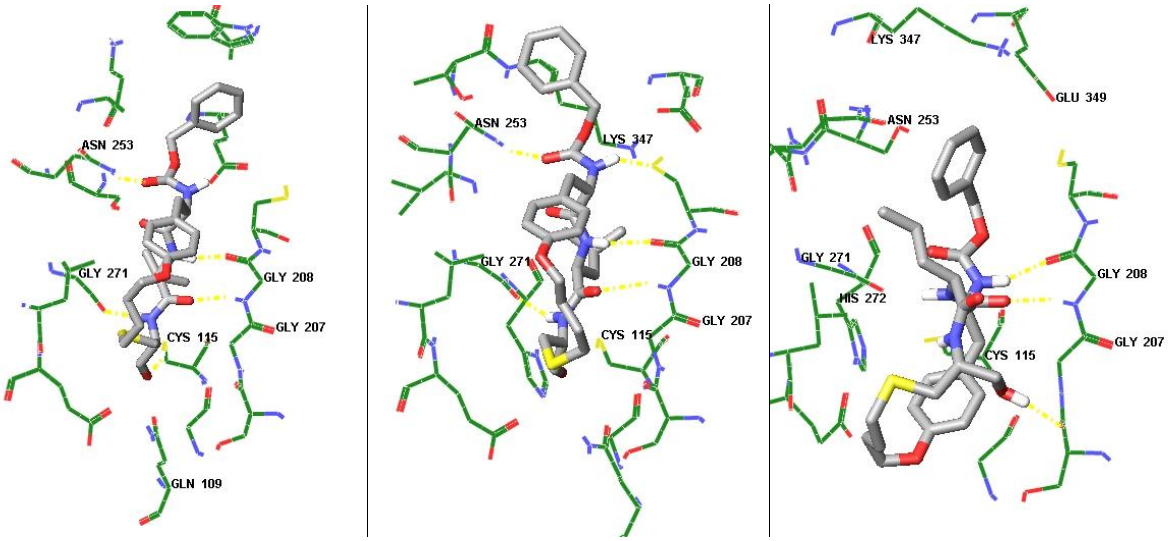


Figure 4.24: Best docked poses of compounds **4.18**, **4.19**, and **4.20** using the InducedFit protocol docked into the calpain model.

Compound **4.19** also docks with the appropriate criteria and has two additional H-bonds. However, the exchange of the valine for a leucine has negatively affected the IC_{50} which is 1010 nM. The effect of the different P_2 side chains on the IC_{50}'s of these two compounds is the opposite to what is seen with compounds **4.12** and **4.13** where the effect of changing from a valine (compound **4.12**) to a leucine (compound **4.13**) had a positive effect on the IC_{50}.

The percentage of conformers within a 12 kJ/mol window found to be in a β-strand for compound **4.20** was calculated at 0% (**Table 4.10**). This means that compound **4.20** did not produce one β-strand conformation in the 12kJ/mol conformational search and strongly favours a turn type conformation. Predictably, when the lowest energy conformer of **4.20** was docked into the enzyme model the docked poses returned had the compound in a turn type conformation which cannot form the three essential H-bonds. The compound had a bad IC_{50} of 28000nM. Because **4.20** does not appear to form a β-strand and favours a turn

conformation it is likely that in a real system it cannot form a β-strand and bind the enzyme favourably.

In summary, compounds **4.9**, **4.12**, **4.15** and **4.18** form a series of Cbz protected aldehyde macrocycles with a valine at the P_2 position. Previous studies had shown that when a macrocycle with a valine at this position was docked into the enzyme model the valine side chain could fit into the deep hydrophobic S_2 subsite favourably and this was the case here with all other criteria met as well. Compounds **4.9** and **4.15** were not synthesised, however, compounds **4.12** and **4.18** were both synthesised and tested and proved to be effective calpain inhibitors.

Compounds **4.10**, **4.13**, **4.16** and **4.19** are a series of compounds based on the compounds **4.9**, **4.12**, **4.15** and **4.18**, respectively, except the valine at the P_2 has been replaced by a leucine. Previous studies showed that a phenylalanine at the P_2 position to be too large for optimal docking, however, the docking of compounds **4.10**, **4.13**, **4.16** and **4.19** provided strong evidence that a leucine at the P_2 position could comfortably dock into the S_2 subsite. In fact compound **4.13** was synthesised and tested to be our most potent compound synthesized in house with an IC_{50} of 30 nM and was shown to retard calcium-induced cataract in cultured ovine lenses. Compound **4.16** had a healthy IC_{50} of 180 nM.

Compounds **4.11**, **4.14**, **4.17** and **4.20** are a series of compounds that are analogous to the compounds **4.10**, **4.13**, **4.16** and **4.19**, respectively, with the aldehyde warhead changed for an alcohol. Because aldehydes are notoriously known to react with other molecules under cellular conditions[10, 25] it was hoped that changing the warhead to an alcohol would still allow the compounds to inhibit calpain by mimicking the tetrahedral intermediate which is

stabilised by H-bonds in the oxyanion hole. If their respective IC_{50}'s were not as good as their aldehyde counterparts, it was hoped that the trade off in potency would overcome the aldehydes innate reactivity to other proteins under cellular conditions.

The modelling results showed that compounds **4.14** and **4.17** were potentially good calpain inhibitors with all the essential criteria for good inhibition met. The compounds were synthesised in house and proved to be moderately good inhibitors. At about 1-2 orders of magnitude less potent than their aldehyde counterparts, **4.13** and **4.16**, they are interesting compounds because unlike the aldehydes they cannot form a covalent bond to the enzyme yet are able to inhibit the enzyme to a moderate degree. They have yet to be tested *in vivo* to see if the trade off in potency is overcome by their proposed lesser tendency to react with other cellular proteins.

Compound **4.11** did meet the main docking criteria, and, the best poses had a high number of 'ugly' internal contacts which suggested that the compound was being slightly 'forced' into the active site of the enzyme model in order to fit and would not be a good inhibitor. It turns out that the compound has a very poor IC_{50} of 31000 nM

The conformational search results of compound **4.20** and the subsequent β-strand analysis showed that it did not form a β-strand. The lowest energy conformer of the compound, which happened to be a turn type conformation, was docked. The predictable outcome of the docking was that it could not dock with the required criteria and the compound docked in the original turn type conformation. Because it could not form a β-strand within the enzyme pocket it was predicted that it would be a poor inhibitor of calpain.

In fact this compound was made and tested to have an IC_{50} of 28000 nM, a very poor inhibitor of calpain.

4.3.6 *In Vivo* study of compound 4.13 (CAT811)

We next chose to investigate the potential of our most potent potential inhibitor, compound **4.13**, to retard the development of calcium-induced cortical cataracts, which have been linked to an overactivity of ovine m-calpain.[24] Lenses from 9-12-month-old lambs were incubated for 48 h in EMEM (Eagle Minimum Essential Medium) culture medium (10 mL) at 37 °C in 5 % CO_2. Compound **4.13** (1 µM) was added to one lens of each of six pairs of sheep lenses in the culture medium. After incubation for 3 h, $CaCl_2$ was added to a final concentration of 5 mM. Intact ovine lenses (*n*=6) were also incubated in EMEM culture medium as a control. After 6 h, all lenses were photographed over a grid, and the opacity was graded by using the software Image-Pro 4.1. Lenses treated with only calcium showed substantial opacity as associated with cataract formation. The presence of compound **4.13** prevented this calcium-induced opacification, and these lenses remained essentially transparent after incubation for 6 h (**Figure 4.25**). Thus, compound **4.13** in the culture medium was able to significantly reduce lens opacity.

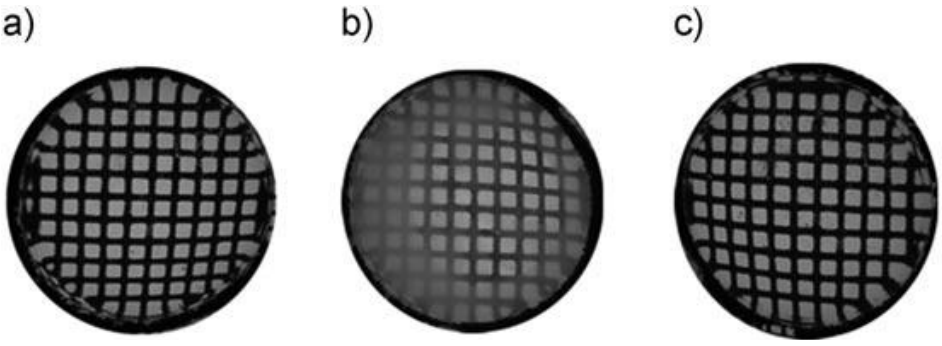


Figure 4.25: Effect of compound **4.13** on calcium-induced cataract formation in cultured ovine lenses. The values given are the mean opacification grading score (SEM: standard error of the mean) in each case and represent the mean opacity of six lenses. Fully opaque lenses have scores of 80-100, whereas transparent lenses

have scores of less than 20. **a)** control: (16.0 ± 2.1); **b)** Ca^{2+} only: (58.8 ± 6.7); **c)** Ca^{2+} with compound **4.13**: (16.7 ± 1.5).

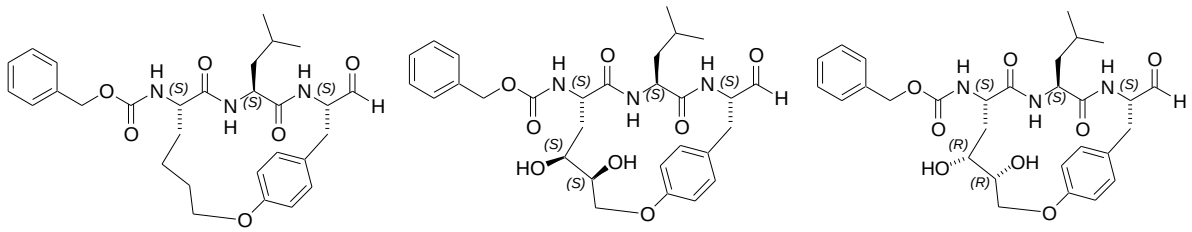
In summary, compound **4.13** (also known as CAT811), the most potent compound *in vitro* with an IC_{50} of 30 nM, was tested in an *in vivo* ovine lens model to see if it could retard the formation of cortical cataracts in ovine lenses. The results of this test showed that compound **4.13** could retard cataract formation in ovine lenses significantly.

4.3.7 Additional modification of macrocyclic cores: docking studies of macrocyclic diols

In the previous section (**Section 4.1.8**) the core macrocycles from **Section 4.1.7** were modified by the substitution of the capping group with Cbz which proved to be a capping group worth investigating further. The studies in **Section 4.1.8** also indicated that leucine at the P_2 position for the Cbz compounds was worthy of further study.

In addition, compound **P8** had been shown (**Section 4.1.7**) to dock into the calpain model using InducedFit with good criteria, therefore, the core of this macrocycle was modified by substiution with Cbz and leucine at the P_2 position (compound **4.21 Figure 4.26**).

The macrocyclic alkene of compound **4.21** was oxidised to diol which gave compounds **4.22** and **4.23** (**Figure 4.26**). It was hoped that docking of these diols would show that the alcohol groups could protrude out of the active site pocket, when the macrocycle is bound in the appropriate way, and into the surrounding hydrophilic environment and that this would be favourable for binding.



4.21 **4.22** **4.23**

Figure 4.26. Chemical structures of the compounds **4.21**, **4.22**, and **4.23**.

Table 4.10: Docking data for best pose[a] (out of possible 20) and inhibitory concentrations (IC_{50}) for compounds **4.21**-**4.23**

Compound	Glide Score [b]	emodel Score [c]	Essential H bonds	Warhead Distance Å	Internal contacts			IC_{50} (nM) m-calpain
					Good	Bad	Ugly	
4.21	-9.2	-73.6	3	3.9	309	17	0	
4.22	-8.1	-74.4	3	4.3	325	19	1	
4.23	-10.3	-68.6	3	3.9	334	22	0	

[a] Best pose chosen by criteria in order of importance; 1 - presence of three essential hydrogen bonds, 2 - warhead within 5Å of nucleophilic Cys115, 3 - low Glide score/emodel score, 4 - lowest number of internal ugly contacts.

[b] GlideScore[18] is a scoring function based on ChemScore[19] which is designed to estimate free energy of binding for protein–ligand complex. The function uses simple contact terms to estimate lipophilic and metal–ligand binding contributions, a simple explicit form for hydrogen bonds and a term which penalises flexibility.

[c] Emodel[18] is a model energy score that combines energy grid score, binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy.

Table 4.11: The Boltzmann weight percentage of conformers within a 12kJ window displaying a *β*-strand for compounds **4.21**-**4.23** and the macrocycle ring size for each compound.

Compounds	**1**	**2**	**3**
% *β*-strand	74	46	10
Macrocycle ring size	17	17	17

The three compounds were docked using **Protocol 5** in the **Appendix** and all display the three essential hydrogen bonds from the *β*-strand backbone to Gly208 and Gly271 of the enzyme (**Table 4.10** and **Figures 4.27**, **4.28**, and **4.29**). The distances between the electrophilic carbon of the ligand and the nucleophilic sulphur of the enzymes active site Cys115 are within 5Å. This indicates that they are close enough for nucleophilic attack by the cysteine sulphur to occur.

Compound **4.21** has an additional hydrogen bond from the carbonyl oxygen of the Cbz group to Lys347 NH_3^+ group. Compound **4.22** has an additional two hydrogen bonds, one from the aldehyde oxygen to the Cys115 SH, the other from the ether oxygen of Cbz to the NH_2 of Asn253. The Glide scores are all low ca -9 which indicates strong binding energies for the enzyme ligand complexes.

All three compounds are positioned within the active site with the backbone being in the favoured *β*-strand conformation. The rest of the macrocyclic rings of **4.21**, **4.22** and **4.23** protrude towards the outside of the cavity along with the diol groups of compound **4.22** and **4.23**. The leucine side chain of all three compounds positioned deep in the hydrophobic region of the active site.

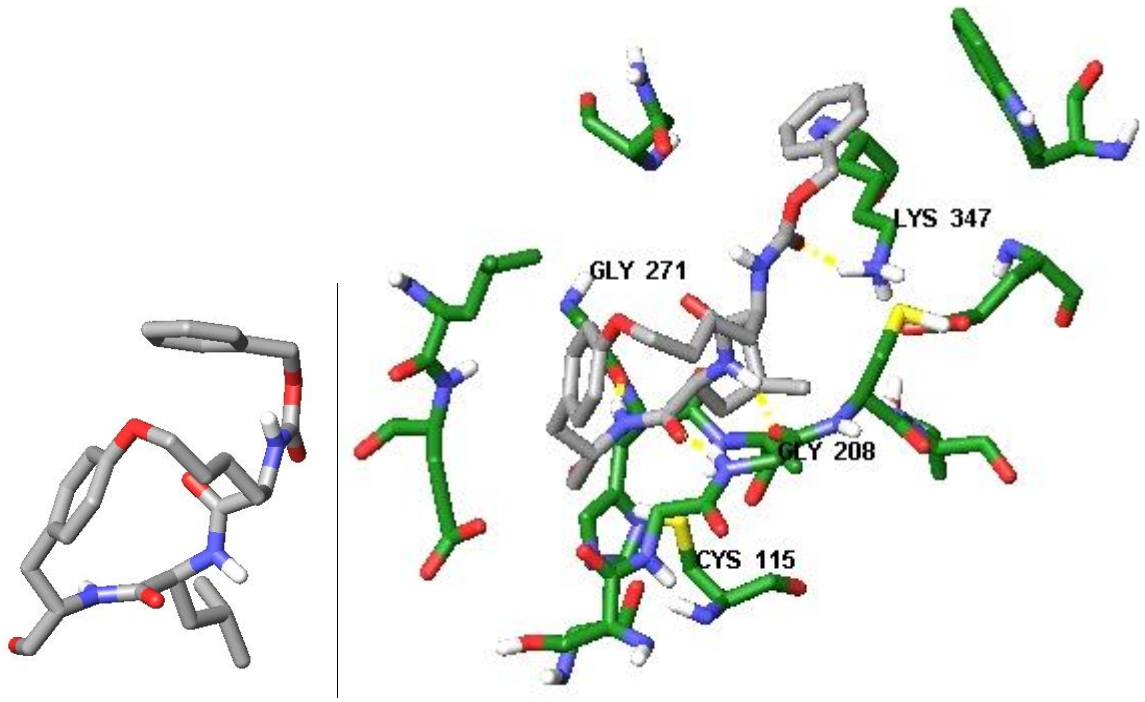


Figure 4.27. Left - lowest energy conformer of compound **4.21**.
Right – best pose of lowest energy conformer of compound **4.21** docked into the enzyme model. Dotted yellow lines show hydrogen bonds between ligand and enzyme.

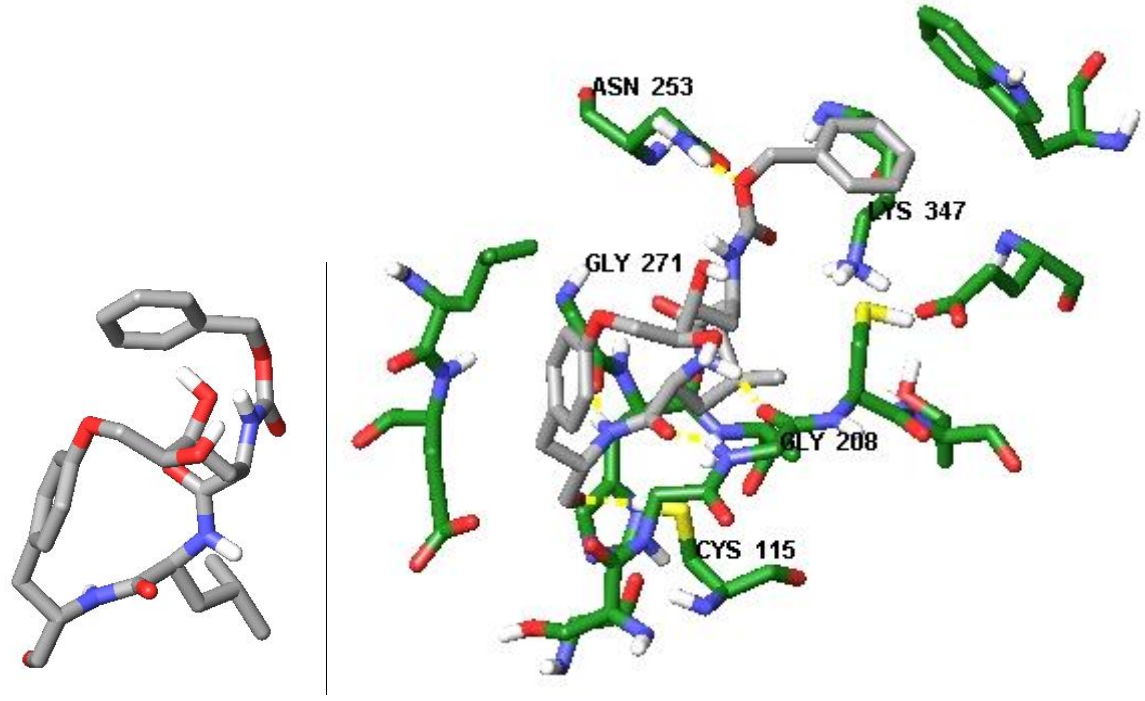


Figure 4.28. Left - lowest energy conformer of compound **4.22** (*S*,*S*-diol).
Right – best pose of lowest energy conformer of compound **4.22** docked into the enzyme model. Dotted yellow lines show hydrogen bonds between ligand and enzyme.

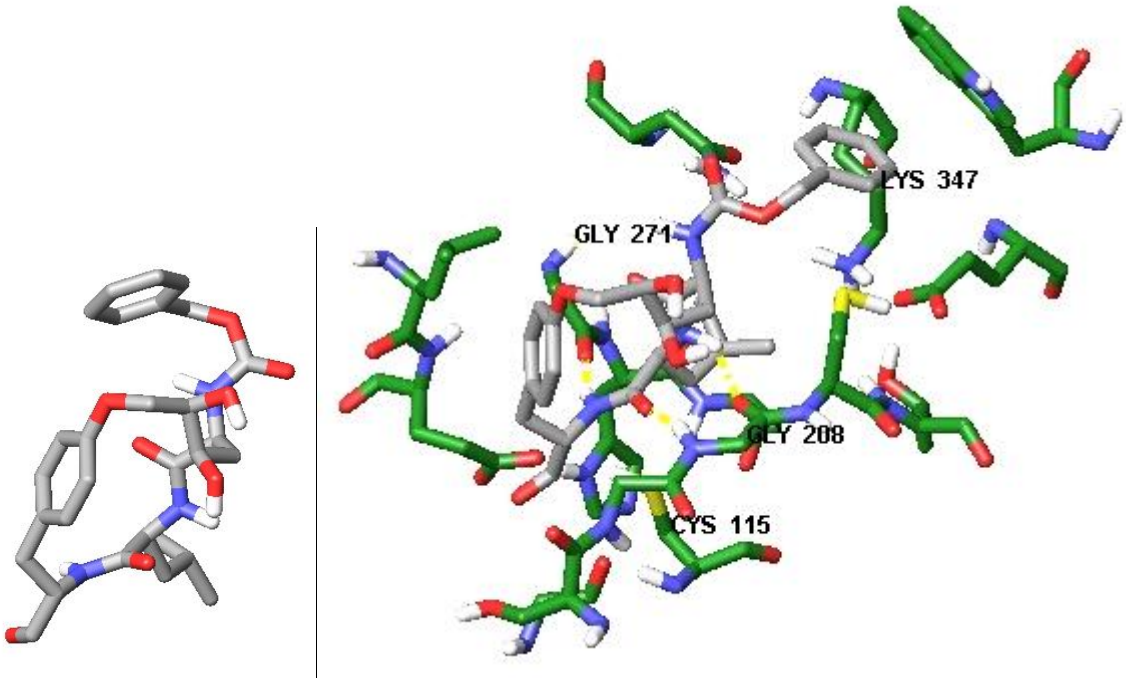


Figure 4.29. Left - lowest energy conformer of compound **4.23** (*R*,*R*-diol).
Right – best pose of lowest energy conformer of compound **4.23** docked into the enzyme model. Dotted yellow lines show hydrogen bonds between ligand and enzyme.

Areas of hydrophobicity and hydrophilicity on the surface of the enzyme crystal structure 1KXR around of the active site are shown in **Figure 4.30**. The top two diagrams show the hydrophobic and hydrophilic areas of 1KXR which has no bound ligand.

When the three compounds are docked (bottom three diagrams of **Figure 4.30**) the hydrophobic leucine side chain fits deep into the hydrophobic pocket of the enzyme. The long hydrophobic area at the top of both top pictures is occupied by the hydrophobic Cbz ring in the bottom three diagrams.

The other large hydrophobic area seen at the right of the top two diagrams (**Figure 4.30**) is an area that changes when each of the three ligands is bound and in a different manner for each ligand. This hydrophobic region sits above the side chain diol regions of compounds **4.22** and **4.23**. The diol regions of these two compounds occupy a proportion of this

hydrophobic surface which is necessarily unfavourable. Compound **4.21**, in the absence of such a diol hydrophobic surface interaction, is much more favourably bound and may therefore be a better inhibitor than compounds **4.22** and **4.23**.

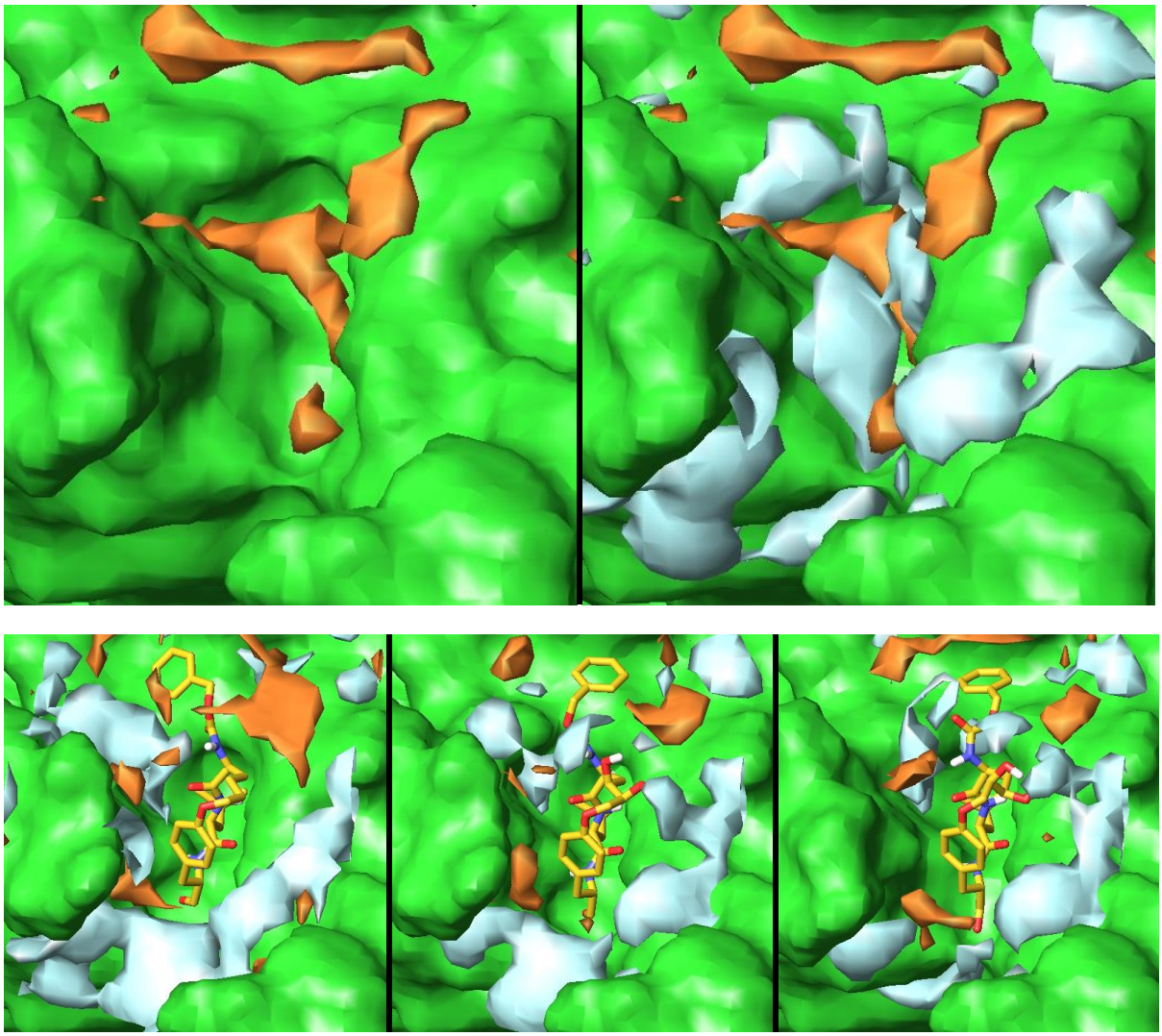


Figure 4.30. Sitemap and surface pictures. Top left – the active site of 1KXR showing area of hydrophobic surface in copper. Top right – the active site of 1KXR showing surface areas that are hydrophobic in a copper colour and the areas that are hydrophilic in a white. Bottom left – best pose of compound **4.21** docked into 1KXR showing hydrophobic and hydrophilic surface areas. Bottom middle – best pose of compound **4.22** docked into 1KXR showing hydrophobic and hydrophilic surface areas. Bottom right – best pose of compound **4.23** docked into 1KXR showing hydrophobic and hydrophilic surface areas.

The hydrophilic area around the opening to the active site pocket in the top right diagram of **Figure 4.28** is due to the hydrophilic side chains of Cys 115, Ser 206, Thr 210, Ser 251, Asn 253, Lys 347, and Glu 349. The backbone NH and CO groups of Gly 207, Gly 208, and

Gly 271, which form the three essential hydrogen bonds between the enzyme and the ligands, also contribute to this hydrophilic area. The formation of these hydrogen bonds is therefore favourable for binding of the three ligands.

In conclusion, all three compounds show the potential to be effective inhibitors of calpain due to their good Glide scores, formation of the three essential hydrogen bonds, and the close proximity of the warhead (aldehyde) to the nucleophilic sulphur of the enzyme.

At the time of writing this thesis we have not received any test results from Joanne Duncan the PhD student who prepared these compounds.

4.4 Conclusion

Cyclization of an inhibitor can reduce their entropic flexibility and constrain the molecule to have propensity for a bioactive conformation. In this case it was preferable to lock the potential inhibitors into a β-strand, a known bioactive conformation for substrates of proteases.

Our first attempt to design cyclic analogues of SJA-6017 was completed before molecular modelling was used as a tool in the design of calpain inhibitors and before any suitable crystal structures of the enzymes were available to our group. Postdoctoral fellow Dr S Miyamoto carried out the first synthetic work on cyclic inhibitors in our research group which was key to establishing ring closing metathesis as a suitable but expensive synthetic strategy, The 8-membered cyclic analogues of SJA-6017 were subsequently shown by

molecular modelling to be non- β-strand mimics and therefore it is not surprising that they did not inhibit m-calpain *in vitro*.

It became increasingly apparent through molecular modelling and literature studies that the β-strand was critical for protease recognition, and therefore inhibition, and that it would be advantageous to design cyclic inhibitors that are locked into this bioactive conformation. This drove our research effort to the design of macrocyclic inhibitors that could mimic a β-strand conformation.

A 288 compound library of possible macrocyclic tri-peptide compounds was proposed with considerable input from Dr Steve Aitken. A selected sample of ten of these compounds was docked using Glide into the enzyme model and this provided evidence that some of these ten compounds could dock into the active site in a fashion that indicated their potential as good inhibitors. It was noticed that some of the compounds could not dock because of turn-type starting conformations which was rectified by re-docking with each of these compounds starting from the lowest energy β-strand conformation. However, some would still not dock with poses that met our criteria of having the three essential H-bonds, close warhead-nucleophile distance, and low Glide and Emodel scores.

Docking studies were then performed with the InducedFit Protocol, which takes up more computer time but was thought to be necessary because of the larger size of the macrocycles compared with the acyclic di-peptides we had previously studied. The InducedFit Protocol allows the active site to make small adjustments with respect to each docked ligand thereby simulating what effectively happens in a real enzyme. InducedFit docking showed that compounds **A1**, **A3**, **A4**, **B1** and **B3** could dock into the enzyme model with all the criteria

met. With this evidence compounds **A1** and **B1** were synthesized and tested *in vitro* and proved to inhibit m-calpain with IC_{50}'s of 3710 and 280 nM, respectively. This was sound evidence that the modelling and docking experiments were providing relevant information to guide the synthetic chemists to synthesizing appropriate compounds from the macrocyclic library.

The ensembles of low energy conformers of the 288 compounds were mixtures of *β*-strands and turn-type conformation. It was thought that by determining the Boltzmann weighted percentage of *β*-strands in each ensemble we could choose appropriate compounds with a propensity of adopting a *β*-strand conformation to model/dock. A program was written that could determine the Boltzmann weighted percentage of *β*-strand in each ensemble.

Compounds in the library that displayed a >90% preference for a *β*-strand (**A1**, **A2**, **H1**, **I9**, **P7**, and **P8**) were then docked using the InducedFit Protocol (compounds A1 and A2 were docked previously). Compounds **A1**, **H1**, **I9**, **P7**, and **P8** showed good docking characteristics. The best docked compound (**H1**) was synthesized and tested to have a moderate IC_{50} of 2400 nM. This again proved that the docking studies were a valuable tool in selection of potential inhibitors.

Although the macrocycles synthesized and tested to date (**A1**, **B1**, and **H1**) were shown to inhibit m-calpain *in vitro*, they were not particularly potent, with the possible exception of **B1** which was a reasonably good inhibitor. To investigate if we could increase the potency of the macrocycles we decided to replace the 4-fluorobenzyl sulphonamide capping group with a Cbz group to the core structures of compounds **A1**, **B1**, **C1**, and **H1**. The cores of **A1**, **B1**, and **H1** were chosen as these three had proven to be moderate inhibitors of m-calpain. The

core of **C1** was included as the previous docking studies had been inconclusive and it was an analogue of **A1** and **B1** with a larger macrocyclic ring.

These four core structures with Cbz (compounds **4.9**, **4.12**, **4.15**, and **4.18**) were also modified to have a leucine (replacing valine) at the P_2 position to give compounds **4.10**, **4.13**, **4.16**, and **4.19**. Compounds **4.10**, **4.13**, **4.16**, and **4.19** were then modified by replacing the aldehyde warhead with an alcohol to give compounds **4.11**, **4.14**, **4.17**, and **4.20**. These twelve compounds were docked into the enzyme model with the InducedFit Protocol. All the compounds with the exception of **4.20** docked in the appropriate manner. Compound **4.20** was not expected to dock appropriately as the Boltzmann weighted percentage of conformers in a β-strand for this compound was calculated at 0% meaning it could not without a considerable energy cost adopt this bioactive conformation.

All the compounds except **4.9** and **4.15** were synthesized and tested including **4.20** (**4.20** was synthesized as a curiosity to see if the modelling was correct in its assumption that it would not be a good inhibitor of m-calpain). The modelling proved correct and **4.20** had a poor IC_{50} of 28000 nM. All the other tested compounds except **4.11** proved to be at least moderately good inhibitors with **4.12** and **4.13** being excellent inhibitors with IC_{50}'s of 85 and 30 nM, respectively. Again the modelling/docking studies proved to be a valuable tool in deciding which compounds to synthesize.

Compound **4.13** (**Figure 4.31**) the best compound in the i*n vitro* test was also tested *in vivo* to evaluate its potential to retard the development of calcium-induced cortical cataracts in sheep lenses. The result proved that **4.13** could significantly retard cataract formation.

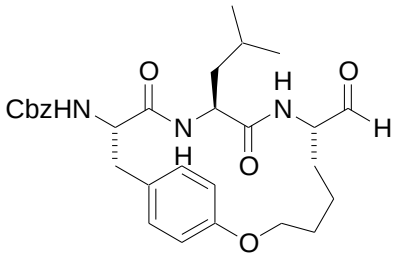


Figure 4.31: Structure of compound **4.13** (also known as Cat 811)

Additional modification of the core structure of compound **P8** was investigated. Cbz had been proven to be a better capping group than 4-fluorobenzyl sulphonamide in the previous study such that **4.13** and similar compounds were better inhibitors than their 4-fluorobenzyl sulphonamide analogues. The core structure of **P8** with a Cbz group, compound **4.21**, was proposed. The alkene of **4.21** was then oxidised to give the diol **4.22** and **4.23**. It was expected that the diol groups would extend out of the active site pocket if the compounds docked appropriately and interact with the hydrophilic environment surrounding the active site. The docking experiments showed that compounds **4.21**, **4.22**, and **4.23** could dock with the key criteria. However, **4.22** and **4.23** docked with their diol groups occupying a normally hydrophobic area to the side of the active site pocket which could be unfavourable for binding, although all three compounds had very good Glide and Emodel scores. At the time of writing the compounds have been prepared but have not yet been tested *in vitro*.

The docking of macrocyclic compounds by the InducedFit Protocol has proven to be a valuable tool in aiding the selection of compounds, from the 288 compound library, to synthesize. It also provided ideas for modification of the core macrocycles with the aim of designing more potent inhibitors of m-calpain.

Despite the model being based on a μ-calpain construct (1KXR) and our ultimate goal being the design of inhibitors of the cataract causing m-calpain, the model has been a success. An m-calpain model would have been a more appropriate model but this was not possible as there is no X-ray crystal structure of an active m-calpain. However, μ-calpain and m-calpain are very similar in their active site structure and most inhibitors of one form inhibit the other and this warranted the use of the μ-calpain construct as a model to perform docking experiments. Molecular modelling cannot be absolutely definitive because nature has too many variables but is a valuable tool in rational drug design and has the potential to get even more effective as appropriate flexibility is built into the enzyme model and computers and algorithms increase speed of the computations necessary.

Future work would be focused on further docking experiments using the macrocyclic library of compounds to find more core structures that show potential as calpain inhibitors. These would be used to design compounds with modified capping groups, warheads, and P_2 side chains. I would also use the X-ray structures of μ-calpain constructs that have inhibitors bound in their active sites to develop a better model, one that could better predict the potential of compounds inhibitory qualities and differential selectivity between the calpains.

4.5 References

1. Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; James, M. N. G., Lowering the Entropic Barrier for Binding Conformationally Flexible Inhibitors to Enzymes. *Biochemistry* **1998,** *37*, 16839–16845.
2. Reid, R. C.; Pattenden, L. K.; Tyndall, J. D. A.; Martin, J. L.; Walsh, T.; Fairlie, D. P., Countering Cooperative Effects in Protease Inhibitors Using Constrained ß-Strand-Mimicking Templates in Focused Combinatorial Libraries. *J. Med. Chem.* **2004,** *47* 1641-1651.
3. Cherney, R. J.; Wang, L.; Meyer, D. T.; Xue, C.-B.; Wasserman, Z.; Hardman, K. D.; Welch, P. K.; Covington, M. B.; R. A. Copeland; Arner, E. C.; DeGrado, W. F.; Decicco, C. P., Macrocyclic Amino Carboxylates as Selective MMP-8 Inhibitors. *J. Med. Chem.* **1998,** *41*, 1749-1751.
4. Lamarre, D.; Anderson, P. C.; Bailey, M.; Beaulieu, P.; Bolger, G.; Bonneau, P.; Bos, M.; Cameron, D. R.; Cartier, M.; Cordingley, M. G.; Faucher, A.; Goudreau, N.; Kawai, S. H.; Kukolj, G.; Lagace, L.; LaPlante, S. R.; Narjes, H.; Poupart, M.; Rancourt, J.; Sentjens, R. E.; George, R. S.; Simoneau, B.; Steinmann, G.; D. Thibeault; Tsantrizos, Y. S.; Weldon, S. M.; Yong, C.; Llinas-Brunet, M., An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **2003,** *426*, 186-189.
5. Shi, Z.-D.; Lee, K.; Liu, H.; Zhang, M.; Roberts, L. R.; Worthy, K. M.; Fivash, M. J.; Fisher, R. J.; Yang, D.; Burke, T. R., A novel macrocyclic tetrapeptide mimetic that exhibits low-picomolar Grb2 SH2 domain-binding affinity. *Biochem. Biophys. Res. Commun.* **2003,** *310*, 378-383.

6. Fairlie, D. P.; Tyndall, J. D. A.; Reid, R. C.; Wong, A. K.; Abbenante, G.; Scanlon, M. J.; March, D. R.; Bergman, D. A.; Chai, C. L. L.; Burkett, B. A., Conformational Selection of Inhibitors and Substrates by Proteolytic Enzymes: Implications for Drug Design and Polypeptide Processing. *J. Med. Chem.* **2000,** *43* 1271-1281.

7. Glenn, P.; Pattenden, L. K.; Reid, R. C.; Tyssen, D. P.; Tyndall, J. D. A.; Birch, C. J.; Fairlie, D. P., ß-Strand Mimicking Macrocyclic Amino Acids: Templates for Protease Inhibitors with Antiviral Activity. *J. Med. Chem.* **2002,** *45* 371-381.

8. Lucke, A. J.; Tyndall, J. D. A.; Fairlie, D. P., Designing Supramolecular Structures from models of Cyclic Peptide Scaffolds with Heterocyclic Constraints. *Journal of Molecular Graphics and Modelling* **2003,** *21*, 341-355.

9. Fukiage, C.; Azuma, M.; Nakamura, Y.; Tamada, Y.; Nakamura, M.; Shearer, T. R., SJA6017, a newly synthesized peptide aldehyde inhibitor of calpain: amelioration of cataract in cultured rat lenses. *Biochimica et Biophysica Acta, Molecular Basis of Disease* **1997,** *1361*, 304-312.

10. Inoue, J.; Nakamura, M.; Cui, Y.; Sakai, Y.; Sakai, O.; Hill, J. R.; Wang, K. K. W.; Yuen, P., Structure-Activity Relationship and Drug Profile of N-(4-Fluorophenylsulfonyl)- L valyl-L-leucinal (SJA6017) as a Potent Calpain Inhibitor. *J. Med. Chem.* **2003,** *46*, 868-871.

11. Tamada, Y.; Fukiage, C.; Mizutani, K.; Yamaguchi, M.; Nakamura, Y.; Azuma, M.; Shearer, T. R., Calpain inhibitor, SJA6017, reduces the rate of formation of selenite cataract in rats. *Current eye research* **2001,** *22*, 280-5.

12. Biswas, S.; Harris, F.; Singh, J.; Phoenix, D. A., The in vitro retardation of porcine cataractogenesis by the calpain inhibitor, SJA6017. *Molecular and Cellular Biochemistry* **2004,** *261*, 169-73.

13. Creighton, C. J.; Reitz, A. B., Synthesis of an Eight-Membered Cyclic Pseudo-Dipeptide Using Ring Closing Metathesis. *Org Lett* **2001,** *3*, 893-5.

14. Fink, B. E.; Kym, P. R.; Katzenellenbogen, J. A., Design, Synthesis, and Conformational Analysis of a Proposed Type I beta-Turn Mimic. *J. Am. Chem. Soc.* **1998,** *120*, 4334-4344

15. Nadin, A.; Derrer, S.; McGeary, R. P.; Goodman, J. M.; Raithby, P. R.; Holmes, A. B., Seven-Membered Lactams as Constraints for Amide Self-Recognition. *J. Am. Chem. SOC.* **1995,** *117*, 9768-9769

16. Mohamadi, F.; Richards, N. G. I.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, C., MacroModel – An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J Comp Chem* **1990,** *11*, 440-467.

17. Shenkin, P. S.; McDonald, D. Q., Cluster Analysis of Molecular Conformations. *J. Comp. Chem.* **2004,** *15* 899-916.

18. Perrin, B. J.; Huttenlocher, A., Calpain. *Int. J. Biochem Cell Biol.* **2002,** *34*, 722–725.

19. Reverter, D.; Sorimachi, H.; Bode, W., The structure of calcium-free human m-calpain: implications for calcium activation and function. *Trends cardiovasc. med.* **2001,** *11*, 222-9.

20. Cuerrier, D.; Moldoveanu, T.; Inoue, J.; Davies, P. L.; Campbell, R. L., Calpain Inhibition by Alpha-Ketoamide and Cyclic Hemiacetal Inhibitors Revealed by X-ray Crystallography. *Biochemistry* **2006,** *45*, 7446-7452.

21. Moldoveanu, T.; Campbell, R. L.; Cuerrier, D.; Davies, P. L., Crystal Structures of Calpain-E64 and-Leupeptin Inhibitor Complexes Reveal Mobile Loops Gating the Active Site. *J. Mol. Biol.* **2004,** *343* 1313.

22. Qian, J.; Cuerrier, D.; Davies, P. L.; Li, Z.; Powers, J. C.; Campbell, R. L., Cocrystal Structures of Primed Side-Extending r-Ketoamide Inhibitors Reveal Novel Calpain-Inhibitor Aromatic Interactions. *J. Med. Chem.* **2008,** *51*, 5264–5270.

23. Griffiths-Jones, S. R.; Sharman, G. J.; Maynard, A. J.; Searle, M. S., Modulation of Intrinsic f,c Propensities of Amino Acids by Neighbouring Residues in the Coil Regions of Protein Structures: NMR Analysis and Dissection of a b-Hairpin Peptide. *J. Mol. Biol.* **1998,** *284*, 1597-1609.

24. Robertson, L. J. G.; Morton, J. D.; Yamaguchi, J. M.; Bickerstaffe, R.; Shearer, T. R.; Azuma, M., *Invest. Ophthalmol. Vis. Sci.* **2005,** *46*, 4634 - 4640.

25. Donker, I. O., A Survey of Calpain Inhibitors. *Curr. Med. Chem.* **2000,** *7* 1171-1188.

5 Drug penetration, *in vivo* testing, and crystallography

5.1 Construction of a modified Ussing chamber

To test the effectiveness of our inhibitors to reach their target (m-calpain in the eye lens) I designed a modified Ussing chamber in an attempt to measure the diffusion rates (penetration) of our inhibitor across a sheep's cornea. The cornea is a major barrier to delivery of drugs into the inner eye tissues.

The Ussing chamber was originally invented by Hans Ussing to study ion transport through frog skin.[1, 2] It consists of a chamber that holds the tissue and buffer solution, and electrical circuitry that measures resistance, current, and voltage so that the transport of ions across the tissue can be studied. For our purposes the electrical circuitry is not needed.

Using an Ussing chamber borrowed from the biology department and data from similar cornea penetration experiments in the literature,[3] we modified the chamber so that it would hold a sheep's cornea between the chamber's two sides.

The plans for the construction of the modified Ussing chamber are shown in **Figures 5.1-5.3**.

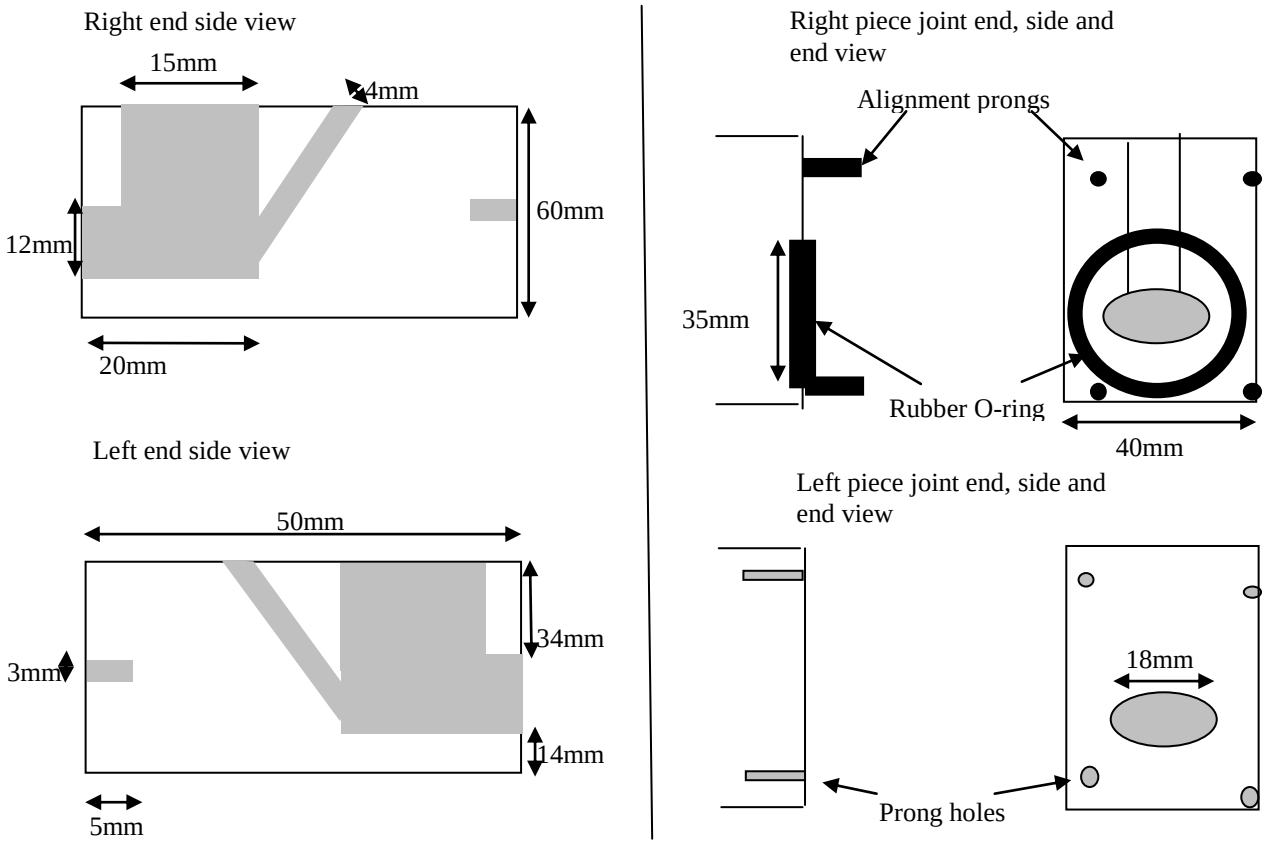


Figure 5.1: (Left) Side view cross sections of modified Ussing chamber showing the dimensions of the separated right and left-handed pieces. Grey areas depict hollowed out parts of the solid perspecs cylinder.
Figure 5.2: (Right) Detail of the joint ends of the modified Ussing chamber. Grey areas depict hollowed out areas of perspecs (chamber connecting holes and prong holes). Black areas depict prongs and the rubber O-ring.

Each side of the chamber holds approximately 5mL and has small gas injection intakes to allow a 95% O_2/5% CO_2 gas to be bubbled through the buffer solutions. The two piece chamber is joined together with four prongs (pins). The tissue (cornea) is placed between the two sides of the chamber so that it covers (seals) the opening between the chambers two sides and the two sides of the chamber are held firmly in place with a specially designed vice grip .

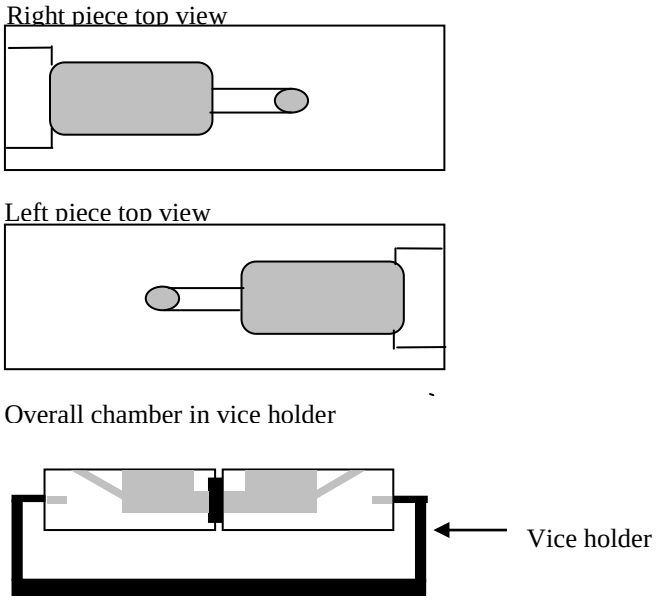


Figure 5.3:. Top views of each end of chamber with the grey areas depicting the openings to the hollowed out areas. Bottom diagram is of whole Ussing chamber setup in vice (black). O-ring is black and the grey areas are showing the hollowed out interior.

5.2 Testing of inhibitors using the Ussing chamber to determine corneal penetration

Two of our compounds (dialdehyde compound **3.26** and triazene compound **3.38a**) and the lead compound **3.10** have been tested using the Ussing chamber to determine how much, if any, of these inhibitors can cross the cornea from the epithelial side (outside) to the endothelial side (inside) and are shown in **Figure 5.4**.

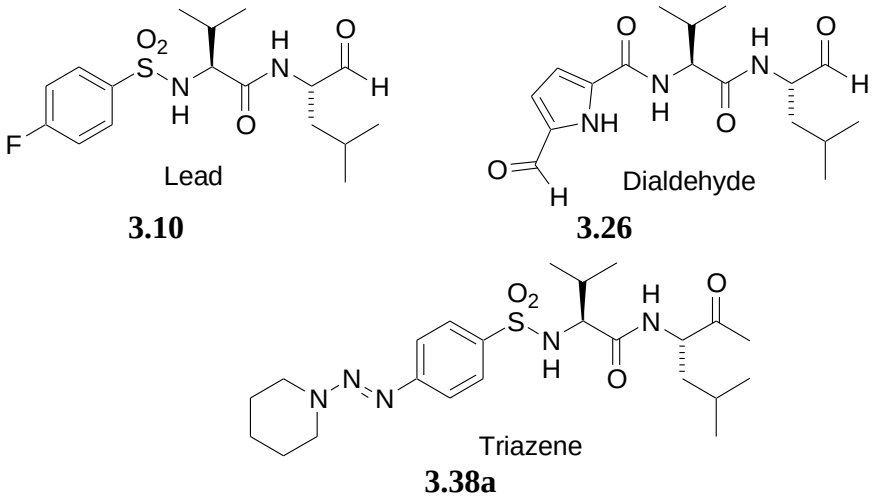


Figure 5.4; Structures used in Ussing chamber experiment. Lead compound **3.10**, dialdehyde compound **3.26** and the triazene compound **3.38a**.

The procedure is as follows:

1. The chamber is partially submerged in a water bath at 35 ºC.

2. The cornea of a freshly harvested sheep's eye is carefully excised and placed so the two sides of the chamber are separated from each other by the cornea.

3. A solution of buffer **A** (5ml) saturated with a test compound at 35 ºC is placed into the epithelial side of the chamber.

   Buffer **A** contains;

   0.1% tween 80
   0.1%$NaH_2PO_4.2H_2O$
   0.9% NaCl
   NaOH to adjust to pH 7.0

4. A solution of buffer **B** (5ml) at 35 ºC is placed into the endothelial side of the chamber.

   Buffer **B** contains;

   0.0132% $CaCl_2$
   0.04 KCl
   0.02% $MgSO_4.2H_2O$
   0.0187% $NaH_2PO_4.2H_2O$
   0.787% NaCl
   0.1% glucose
   NaOH to adjust to pH 7.2

5. A 95% O_2/5% CO_2 gas (carbogen gas) is injected into both sides of the chamber at a rate of 2-3 bubbles per second.

6. The chamber is kept at 35 ºC throughout experiment.

7. 200µl aliquots from each side of the chamber are taken at regular intervals (every 30 minutes) over 3 hours.

8. The aliquots were examined by HPLC to determine the concentration of inhibitor in each aliquot.

9. A plot of concentration of compound (inhibitor) versus time gives a qualitative measure of the diffusion rate of inhibitor through the sheep's cornea.

The buffers are formulated differently to simulate the actual environment of the epithelial and endothelial sides of the cornea in a live sheep.

The aliquots taken at 30min intervals were analysed by HPLC by Seth Jones and Dr Mathew Jones. However, the aliquots taken for all three compounds had no detectable traces of the respective compounds. The compounds were not able to cross the cornea in detectable (by HPLC) amounts over the 3 hour time period.

5.3 *In-vivo* sheep trial results of compound 3.26

Our collaborators at Lincoln University were given our dialdehyde compound **3.26** to conduct *in vivo* sheep trials. A ten week trial commenced in October 2004 using an eye drop formulation of compound **3.26**, which had the following formula;

0.1% (w/w) compound **3.26**
14% EtOH

0.9% sodium chloride
0.3% hydroxypropyl methyl cellulose
0.05% disodium EDTA
0.01% benzalkonium chloride
84.65% MilliQ purified water

The formula of compound **3.26** was applied to the left eye of lambs with developing cataracts, leaving the right eye as an untreated control. The development of cataracts was scored based on the scoring system in **Figure 5.5** by a veterinary ophthalmologist two weeks before the start of the evaluation (t= -14) and three times during the eleven weeks of treatment (t = 10, 37, 67 days). The mean score after each examination is shown in **Figure 5.6**.

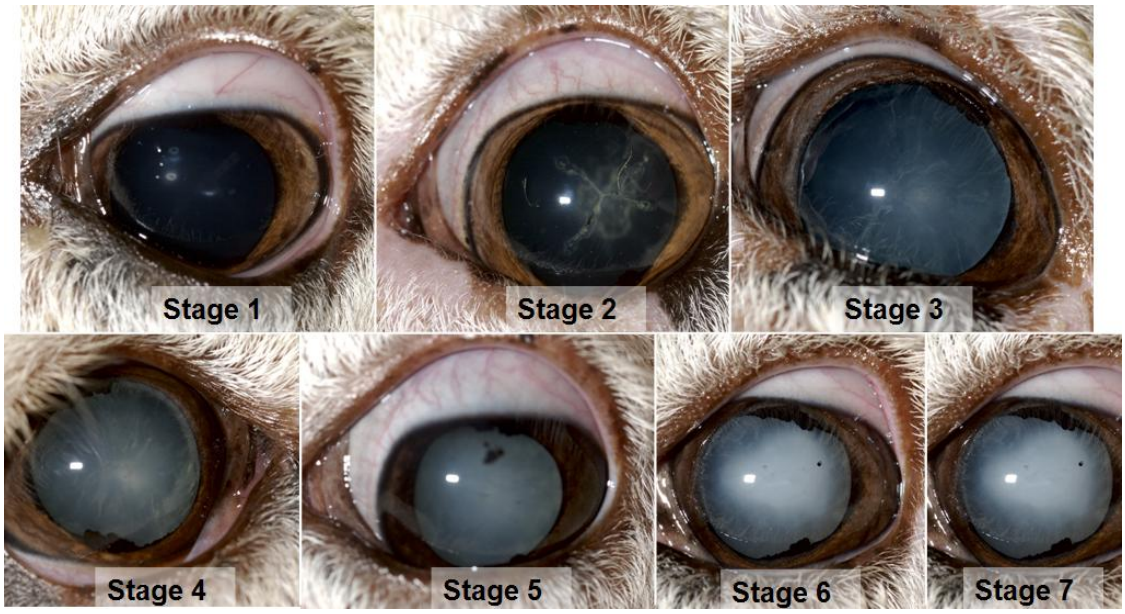


Figure 5.5: One to seven cataract scoring system. **(1)** Anterior suture lines **(2)** Anterior & posterior suture lines **(3)** 0-33% cortical nuclear involvement **(4)** 33-66% cortical nuclear involvement **(5)** Greater than 66% cortical nuclear involvement **(6)** Total immature cataract **(7)** Total mature cataract

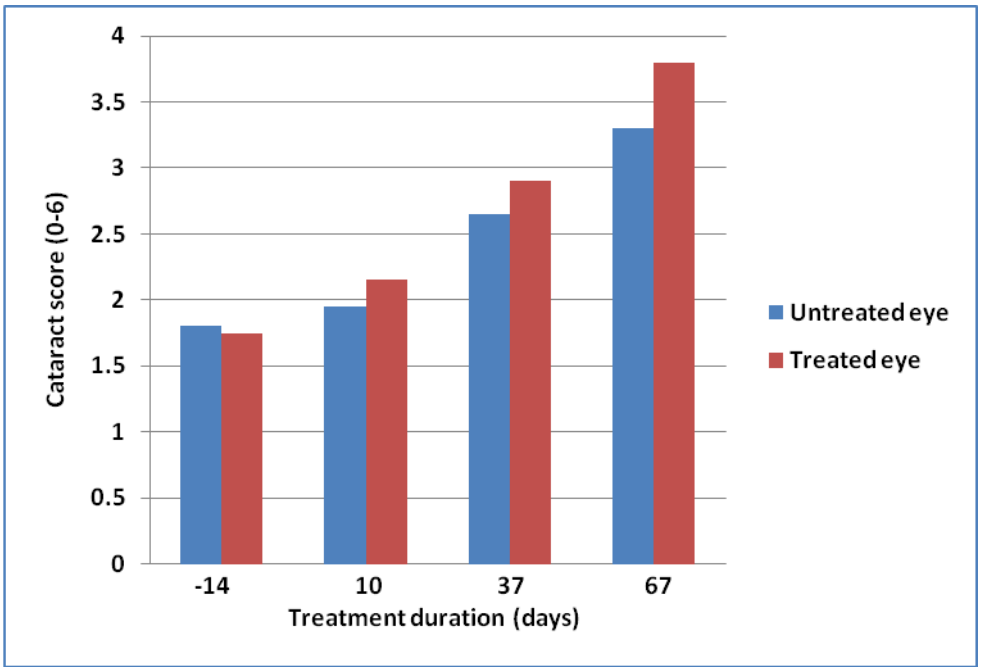


Figure 5.6: A plot of mean cataract scores for untreated and treated eyes

Both eyes had the same cataract scores over the first 37 days. By the end of the 67 day trial the treated eye scored higher than the untreated eye. The trial had shown that the formula containing compound **3.26** did not retard cataract development in the trial conditions. The reason why no retardation was observed was postulated as being inadequate diffusion of the compound across the cornea which was likely due to the use of ethanol in the formulation that caused the eye drop to have a low surface tension. This caused most of the eye drop to be washed off the surface of the eye.

To overcome this problem another trial was performed the following October which used compound **3.26** in an ointment formulation. It was hoped that the ointment would stay on the surface of the eye for a longer period of time in order to increase the amount of drug able to diffuse across the cornea.

The formula was as follows;

1%	compound **3.26**
25%	cetyl stearyl alcohol
35%	lanolin
39%	paraffin oil.

Twenty-four two to three month old lambs were used for the trial. The left eye of each sheep was treated twice daily with twenty five milligrams of ointment containing 0.25 milligrams of compound **3.26**.

During the trial the cataracts progressed rapidly and by the end of the trial over half had mature cataracts (**Figure 5.7**)

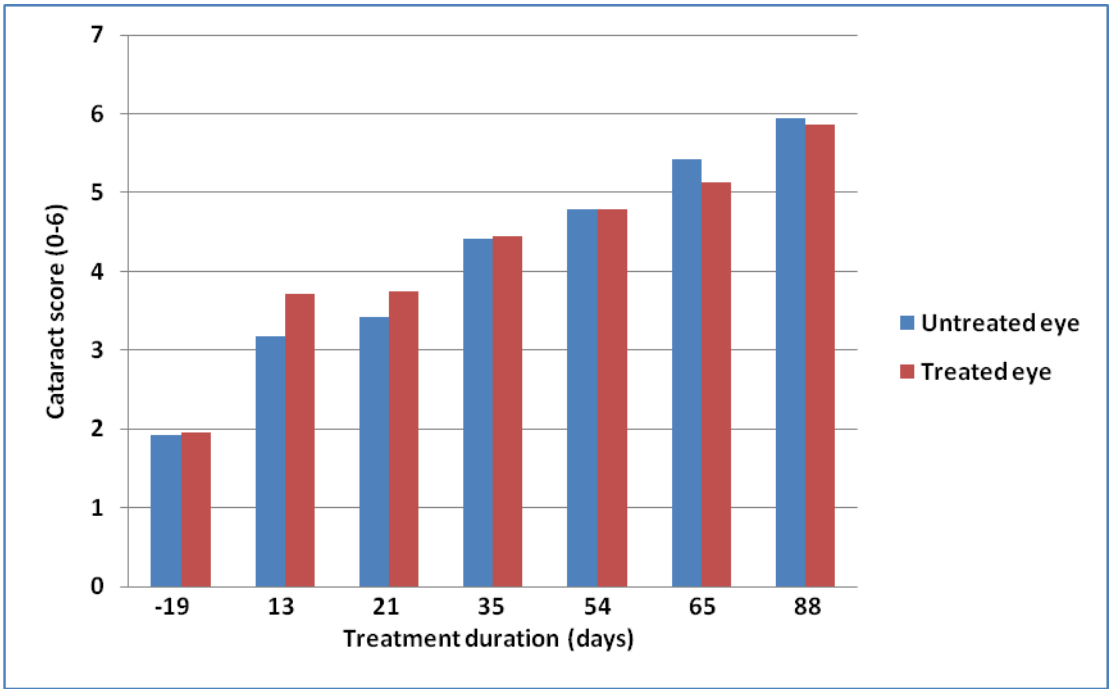


Figure 5.7: Results of the ointment containing compound **5** in the *in-vivo* cataract sheep trial.

The graph in **Figure 5.7** shows that after one month the treated eyes showed significant cataract retardation (significant in a paired t-test at $p < 0.05$).

Compared with the cataract development in the sheep used in this trial, human cataracts develop at a much slower rate, therefore, the significant retardation of the sheep cataracts after one month would be expected to be significantly longer for humans. However, by the end of the trial there was no significant difference between treated and untreated eyes.

5.4 Crystallography

An attempt to co-crystallize papain with a number of our best inhibitors using the hanging drop method was performed. The procedure for this was based on the methods used in J. LaLonde 1998.[4]

The hanging drop procedure for this went as follows:

(i) A reservoir solution was made up consisting of;

0.1M Tris HCl pH 8.5

0.5M Trisodium citrate

20% PEG 600

(ii) Papain purchased from Sigma was centrifuged to remove buffer yielding a 35mg/ml solution in water.

(iii) 0.3 M solutions of the 7 best inhibitors (**3.10**, **3.26**, **3.27**, **3.29**, **3.31**, **3.38a**, and **3.38b**) were made up in DMSO.

(iv) 6 µl solution of papain, 2 µl solution of inhibitor, and 8 µl solution of precipitant (reservoir solution) were added together and placed on a cover slip. This procedure was repeated for each inhibitor in duplicate plus four blanks without an inhibitor.

(v) 1ml solution of reservoir was placed into each well. The cover slips with papain, inhibitor, and precipitant were overturned and sealed on top of the wells with paraffin to form a hanging drop over the reservoir solution.

(vi) The wells were then placed in a refrigerator at 5 ºC.

(vii) Regular checks were made to identify any crystals that may have formed.

This procedure failed to produce any crystals. It is possible that the papain was not pure enough to produce crystals and it may be necessary to further purify the papain by affinity chromatography or dialysis.

5.5 Conclusion

The Ussing chamber experiment did not produce any results because the diffusion of the inhibitors across the sheep corneas was too slow or non-existent over the three hour experiment and consequently there was not enough of each inhibitor on the endothelial side of the chamber to be detected by HPLC.

To overcome this detection problem in the future, the experiment could be run over night for 24 hours to determine if the inhibitors can cross the cornea over a longer period of time. Another possibility would be to label the inhibitors using a radioactive label. Commonly used radioisotopes used in this way are ^{3}H, ^{14}C, ^{15}N, and ^{18}O.[5] An inhibitor synthesized with a radioisotope could then be used in the Ussing chamber diffusion experiment so that the concentration of radioactive inhibitor on the endothelial side could be measured. However, the equipment and resources needed to carry out such an undertaking could be expensive.

The *in vivo* sheep trial showed that drug formulation is a critical aspect for the treatment of cataract because of the fact that the inhibitors are easily washed away by the continuous flushing of the eye by tears as was seen in the first trial using ethanol. The formulation in the next trial proved to be better as the ointment was able to stick to the eye for a longer period of time which increased the likelihood of corneal penetration of the inhibitor. Treatment showed significant results after one month but this was not sustained over the entire trial.

5.6 References

1. Ussing, H. H.; Lind, F., Trapping of 134Cs in frog skin epithelium as a function of shortcircuit current. *Kidney International* **1996,** *49*, 1568-1569.
2. Ussing, H. H.; Zerahn, K., Active transport of sodium as the source of electric current in the short circuited isolated frog skin. *Acta Physiol. Scand.* **1951,** *23*, 110-127.
3. Chung, Y. B.; Han, K.; Nishiura, A.; Lee, V. H. L., Ocular Absorptioin of Pz-Peptide and Its Effect on the Ocular and Systemic Pharmacokinetics of Topically Applied Drugs in the Rabbit. *Pharmaceutical Research* **1998,** *15*, 1882-1887.
4. LaLonde, J. M.; Zhao, B.; Smith, W. W.; Janson, C. A.; DesJarlais, R. L.; Tomaszek, T. A.; Carr, T. J.; Thompson, S. K.; Oh, H.; Yamashita, D. S.; Verber, D. F.; Abdel-Meguid, S. S., Use of Papain as a Model for the Structure-Based Design of Cathepsin K Inhibitors: Crystal Structures of Two Papain-Inhibitor Complexes Demonstrate Binding to S'- Subsites. *J. Med. Chem.* **1998,** *41*, 4567-4576.
5. Rennie, M. J., An introduction to the use of tracers in nutrition and metabolism. *Proceedings of the Nutrition Society* **1999,** *58*, 935-944.

6 Design of antimicrobials and herbicides

6.1 Introduction

Work was also performed examining specific targets in the shikimate pathway a pathway important in plants and microbes, but not observed in the animal kingdom.[1-3] It is therefore a suitable pathway to block when designing a herbicide or antimicrobial since it would be expected to have no effect on humans or animals. This pathway has been successfully targeted by the herbicide glyphosate which binds the sixth enzyme in the shikimate pathway known as 5-enolpyruvylshikimate-3-phosphate-synthase (EPSP synthase).[2]

Another enzyme in the shikimate pathway known as 3-dehydroquinase, which catalyses the third step, is a sensible target for inhibitor action (**Figure 6.1**). What makes it attractive is that two different enzymes have evolved, type I and type II and both catalyze the dehydration of 3-dehydroquinate to 3-dehydroshikimate, but each has a distinct mechanism of action.[4] The type I enzymes catalyze a *cis*-dehydration of 3-dehydroquinate via a covalent imine intermediate, while the type II enzymes catalyze a *trans*-dehydration via an enolate intermediate.[4] Type II is involved in another pathway, known to occur in some organisms such as fungi, called the quinate pathway.[5]

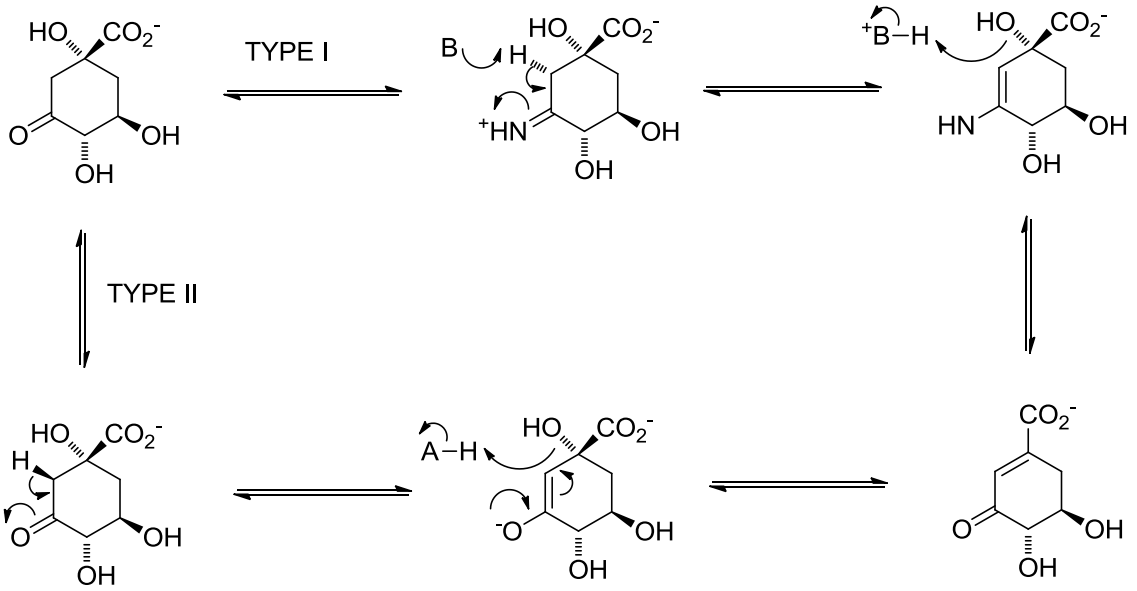


Figure 6.1. Third step in the shikimate pathway by dehydroquinase type I and type II.

As there are two distinct forms of the enzyme there is the opportunity to design inhibitors of one type that do not affect the other. This means that certain organisms that possess one type of enzyme could be selectively targeted.[1] One such example is the infectious bacterium, *Mycobacterium tuberculosis*, which possesses only type II 3-dehydroquinase. Because most (if not all) bacteria associated with human natural flora have only type I, an inhibitor of type II should be an effective antibiotic against tuberculosis, and this is important since it does not upset the natural flora associated with the human body.

There are a number of X-ray structures published including those of the apo-enzymes (without substrates) and those with known inhibitors bound. The most suitable ones investigated to date are PDB codes 1GU1, 1GU0, 1HOR, and 1HOS.[1, 6]

6.2 Modeling studies of dehydroquinase

Five compounds (**6.1-6.4**) were synthesized in house by Dr Mary Gower as analogues of the known dehydroquinase inhibitors **6.5** and **6.6** (**Figure 6.2**). Compounds **6.1-6.4** were run through a conformational search to find low energy conformers using **Protocol 3** in the **Appendix**. The lowest energy conformer of each compound and the X-ray structures of compounds **6.5** and **6.6** were used to initiate Glide docking into a model of the X-ray structure 1HOS. The model was produced using **Protocol 7** in the **Appendix**. Glide docking was performed using **Protocol 8** in the **Appendix**.

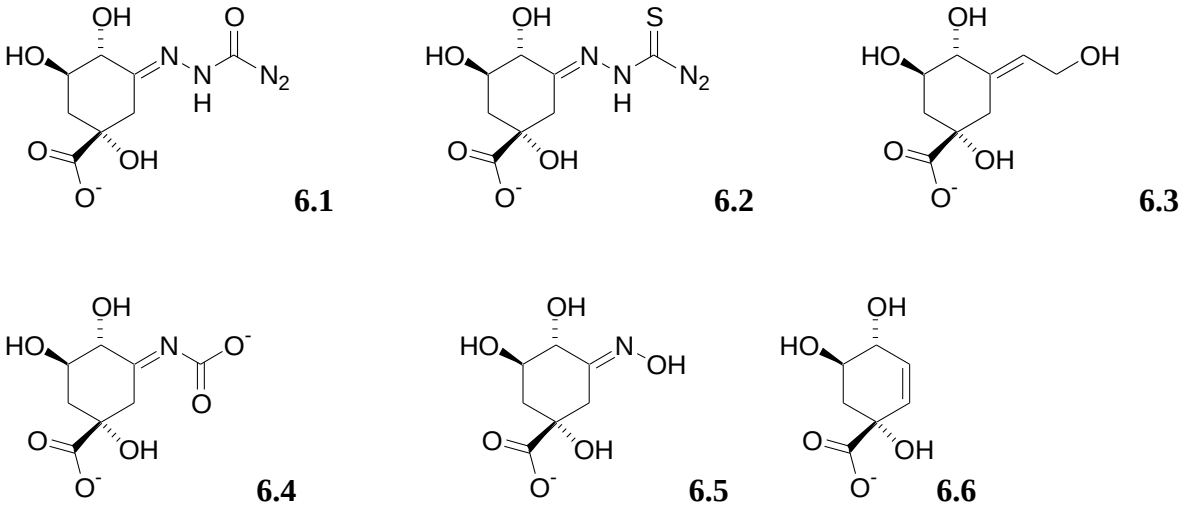


Figure 6.2. Chemdraw structures of compounds **6.1**, **6.2**, **6.3**, **6.4**, **6.5 (**3-hydroxyimino-quinic acid), and **6.6 (**2,3-anhydro-quinic acid).

The X-ray crystal structure of 3-hydroxyimino-quinic acid bound into the active site of dehydroquinase from *Mycobacterium tuberculosis* (MTDHQase) is shown in **Figure 6.3** left (PDB code 1H0S). On the right of **Figure 6.3** is the X-ray crystal structure of 2,3-anhydro-quinic acid bound into the same enzyme (PDB code 1H0R). Both inhibitors form the same eight core H-bonds to the enzyme; one each from His 81, His 101, Ile 102, and Arg 112, and two each from Asn 75 and Ser 103. However, the oxime of 3-hydroxyimino-quinic acid can

form two additional H-bonds to Pro 11 and Arg 19 which are thought to contribute significantly to the binding and is why 3-hydroxyimino-quinic acid is more potent *in vitro* than 2,3-anhydro-quinic acid against dehydroquinase from *M. tuberculosis* (**Table 6.1**).

Table 6.1 Docking data for compounds **6.1**, **6.2**, **6.3**, **6.4**, **6.5**, and **6.6**. All docked with Glide into model of MTDHQase (PDB code 1H0S).

Compound	Glide score	Emodel score	H-bonds	Essential H-bonds out of 5	K_i (μM) MTDHQase
6.1	-6.4	-56.7	7	4	45±5
6.2	-6.0	-54.5	7	4	425±75
6.3	-6.7	-53.5	6	4	1100±100
6.4	-6.3	-44.4	7	3	2000±600
6.5	-7.6	-72.5	10	5	20±2
6.6	-7.5	-60.7	8	5	200±20

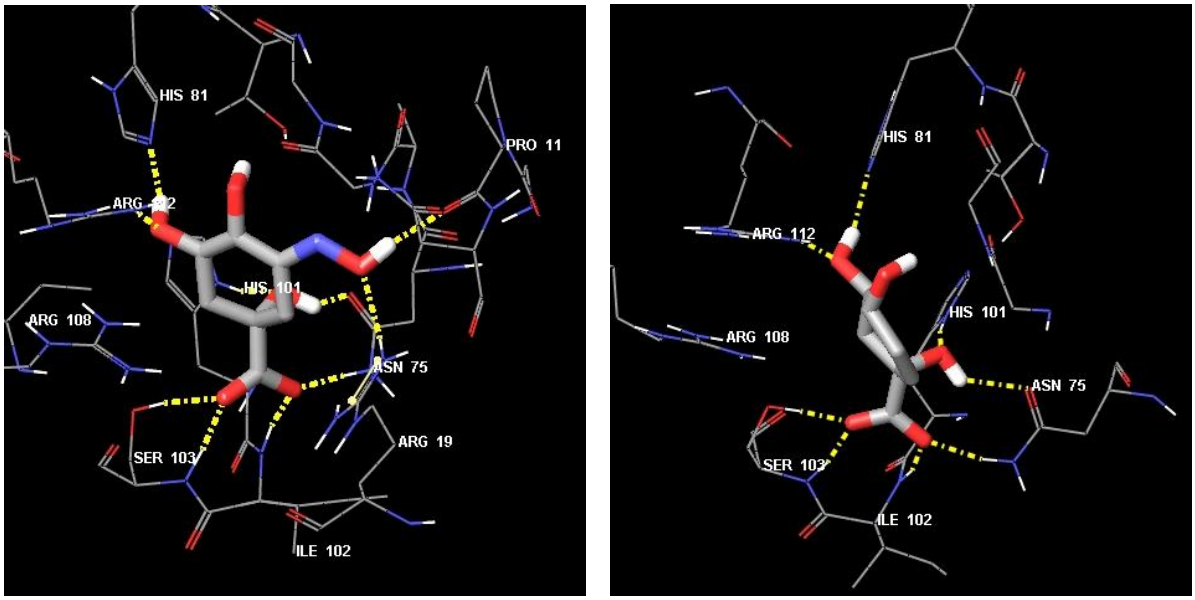


Figure 6.3. Left; X-ray crystal structure of 3-hydroxyimino-quinic acid (compound **6.5**) bound into the active site of MTDHQase (PDB code 1H0S). Right; X-ray crystal structure of 2,3-anhydro-quinic acid (compound **6.6**) bound into the same enzyme (PDB code 1H0R).

The five hydrogen bonds to Ser103, Ile102, and Asn75 were deemed to be essential interactions as these are observed in all known co-crystallised structures with inhibitors of

this type. Consequently, these five hydrogen bonds were used as constraints in the docking of **6.1**-**6.6.**

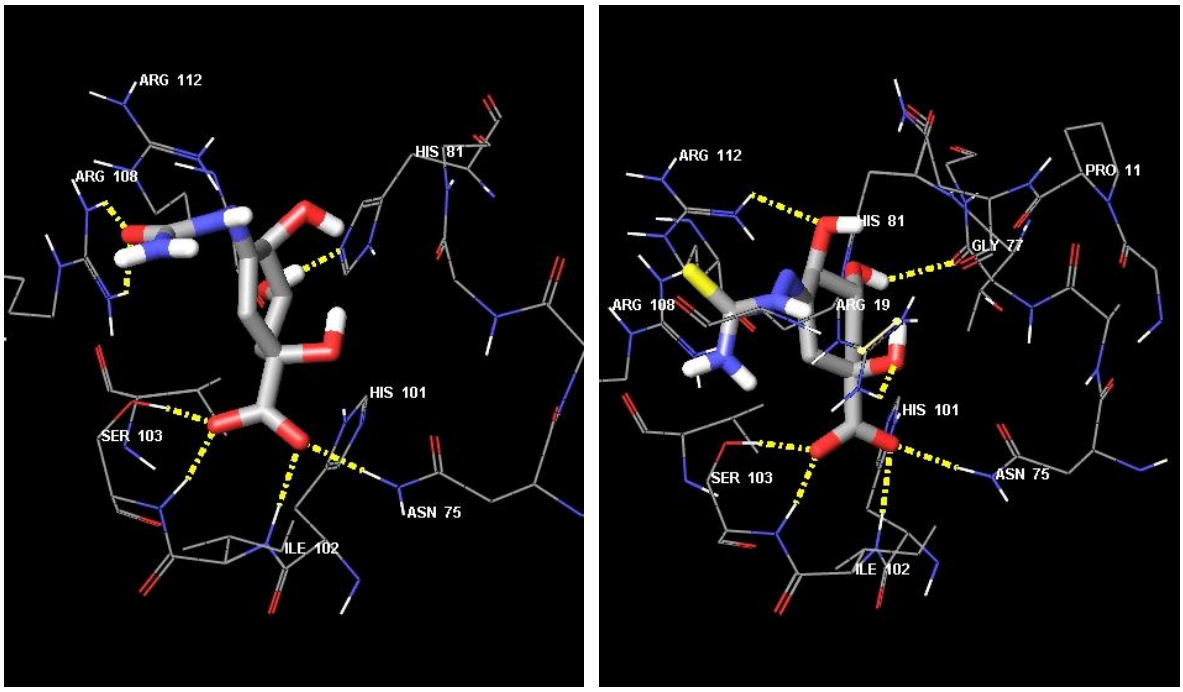


Figure 6.4. Best docked poses of **6.1** (left) and **6.2** (right) into 1H0S (MTDHQase).

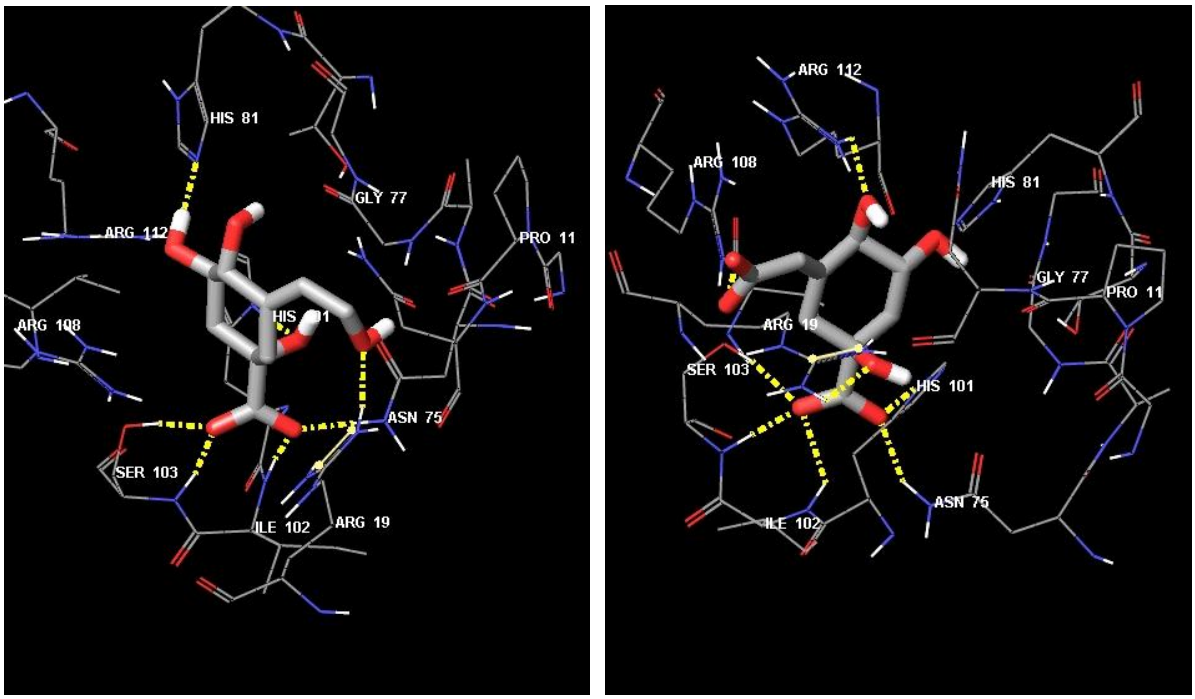
To make certain that the model was able to dock the compounds with good accuracy the compounds **6.5** and **6.6** were docked back into the model to give a 'baseline' to work with in terms of the Glide and Emodel scores.

Both compounds **6.5** and **6.6** dock into the model in almost the exact way as they are seen in their respective co-crystallised structures (**Figure 6.3**). The Glide scores and Emodel scores are both good and reflect the fact that compound **6.5** is more potent *in vitro* than compound **6.6** (**Table 6.1**) as it has a better Glide and Emodel score and it has two extra H-bonds.

Compound **6.1** can form seven H-bonds to the enzyme active site (**Figure 6.4**) and it has four out of the five essential H-bonds used as constraints in the docking. The other H-bonds are

two from the semicarbazone to Arg 108 and one from the 5-hydroxyl to His 81. The H-bond to His 81 is also observed in the crystal structures with compounds **6.5** and **6.6** bound (**Figure 6.3**). The Glide score and Emodel score of -6.4 and -56.7, respectively, indicates good binding. *In vitro* tests showed that compound **6.1** was a potent inhibitor of type II 3-dehydroquinase with a K_i of 45±5 µM (**Table 6.1**). This is almost as potent as the 3-hydroxyimino-quinic acid which had a K_i of 20±2 µM and is more potent than the 2,3-anhydro-quinic acid with a K_i of 200±20 µM (**Table 6.1**).

Figure 6.4 shows that the thiosemicarbazone of **6.2** cannot form H-bonding to the side chain of Arg 108 as is seen with the semicarbazone of **6.1**. However, it does form four out of the five essential H-bonds used as constraints in the docking and it has another three H-bonds from the three hydroxyl groups, one to Gly 77, one to Arg 19, and the other to Arg 112. Compound **6.2** has a slightly inferior Glide and Emodel score compared with compound **6.1** and may indicate why compound **6.2** has a less potent K_i of 425±75 µM (**Table 6.1**).



6.3 **6.4**

Figure 6.5. Best docked pose of **6.3** (left) and **6.4** (right) into 1H0S (MTDHQase).

The hydroxyl group on the alkene of **6.3** can form a H-bond to the side chain of Arg 19 and the carboxyl group on the alkene of **6.4** forms a H-bond to the side chain of Arg 108 (**Figure 6.5**). Compound **6.3** forms four out of five essential H-bonds but only six in total, while **6.4** can only form three essential H-bonds and seven in total. Compound **6.4** also has a much lower Emodel score compared with the other compounds. The loss of H-bonding for both **6.3** and **6.4** and the lower Emodel scores is possibly why these compounds are significantly less potent than compound **6.1**, the hydroxyimino-quinic acid **6.6**, and the 2,3-anhydro-quinic acid **6.5** with a K_i of 1100±100 µM for **6.3**, and 2000±600 µM for **6.4** (**Table 6.1**).

6.3 Conclusion

Four compounds (**6.1**-**6.4**) synthesized in house by Dr. Mary Gower were docked into the dehydroquinase model which was based on the X-ray crystal structure of type II dehydroquinase from *M. Tuberculosis* (1H0S). These were compared with two known inhibitors (**6.5** and **6.6**) which have published X-ray co-crystal structures with the enzyme.

The compounds were also tested *in vitro* to give K_i values which gave data that was used to compare docking results. The results of the docking study showed that the known inhibitors (**6.5** and **6.6**) could re-dock back into their respective X-ray crystal structures in exactly the same way as they are seen in these structures and had excellent Glide and Emodel scores which gave a 'baseline' to compare with the docking data from the in house compounds (**6.1**, **6.2**, **6.3**, and **6.4**). These four compounds docked in a similar way to the two known inhibitors. Compound **6.1** had a Glide and Emodel score that was comparable to the known inhibitors, it formed a total of seven H-bonds with four H-bonds being part of the essential five used as constraints in the docking. **6.1** proved to be the most potent of the in house compounds with a K_i of 45±5 μM which is almost as potent as compound **6.5**.

Compounds **6.2**, **6.3**, and **6.4** are less potent inhibitors. **6.2** is an order of magnitude less potent than **6.1** and has a slightly poorer Glide and Emodel score. **6.3** forms the least number of H-bonds to the active site and this may be why it has a poor K_i of 1100±100 μM. **6.4** is the least potent and this is reflected by it having the lowest Emodel score and only forming three out of the five essential H-bonds.

6.4 References

1. Roszak, A. W.; Robinson, D. A.; Krell, T.; Hunter, I. S.; Fredrickson, M.; Abell, C.; Coggins, J. R.; Lapthorn, A. J., The structure and mechanism of the type II dehydroquinase from Streptomyces coelicolor. *Structure* **2002,** *10*, 493.
2. Roberts, F.; Roberts, C.; Johnson, J. J.; Kylek, D. E.; Krell, T.; Coggins, J. R.; Coombs, G. H.; Milhousk, W. K.; Tzipori, S.; FergusonI, D. J. P.; Chakrabarti, D.; McLeod, R., Evidence for the shikimate pathway in apicomplexan parasites. *Nature* **1998,** *393*, 801-805.
3. Robinson, D. A.; Stewart, K. A.; Price, N. C.; Chalk, P. A.; Coggins, J. R.; Lapthorn, A. J., Crystal Structures of *Helicobacter pylori* Type II Dehydroquinase Inhibitor Complexes: New Directions for Inhibitor Design. *J. Med. Chem.* **2006,** *49*, 1282-1290.
4. Gourley, D. G.; Shrive, A. K.; Polikarpov, I.; Krell, T.; Coggins, J. R.; Hawkins, A. R.; Isaacs, N. W.; Sawyer, L., The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. *Nature Structural Biology* **1999,** *6*, 521-525.
5. Frederickson, M.; Parker, E. J.; Hawkins, A. R.; Coggins, J. R.; Abell, C., Selective Inhibition of Type II Dehydroquinases. *J. Org. Chem.* **1999,** *64*, 2612-2613.
6. Robinson, D. A.; Roszak, A. W.; Frederickson, M.; Abell, C.; Coggins, J. R.; Lapthorn, A. J., Structural Basis for Selectivity of Oxime Based Inhibitors Towards Type II Dehydroquinase from Mycobacterium Tuberculosis. To be Published.

7 Molecular modelling to determine the absolute configuration of chrysosporide

The following is data published in the Journal of Natural Products;
Mitova, M. I.; Stuart, B. G.; Cao, G. H.; Blunt, J. W.; Cole, A. L. J.; Munro, M. H. G., Chrysosporide, a Cyclic Pentapeptide from a New Zealand Sample of the Fungus *Sepedonium chrysospermum*. *J. Nat. Prod.* **2006,** *69,* 1481-1484.

7.1 Introduction

A new cyclic pentapeptide, chrysosporide (**7.1**), was isolated from a New Zealand sample of the mycoparasitic fungus Sepedonium chrysospermum[1] by bioactivity-guided fractionation. A planar structure was deduced by detailed spectroscopic analysis, and the absolute configurations of the amino acid residues were defined by Marfey's method.[2] As both enantiomers of Leu occurred in chrysosporide, molecular mechanics calculations were applied to the analysis to distinguish between the possible structural isomers. Only the lowest energy conformers of the cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu) isomer were in agreement with the observed NOEs, suggesting that this was the most probable amino acid sequence for chrysosporide (**7.1**).[3]

Mycoparasitic fungi are a diverse and prolific source of compounds with potential therapeutic value.[3] In our continuing search for new, bioactive metabolites from New Zealand fungi a strain of Sepedonium chrysospermum was investigated, as the culture extracts were cytotoxic against the P388 murine leukemia cell line.[4]

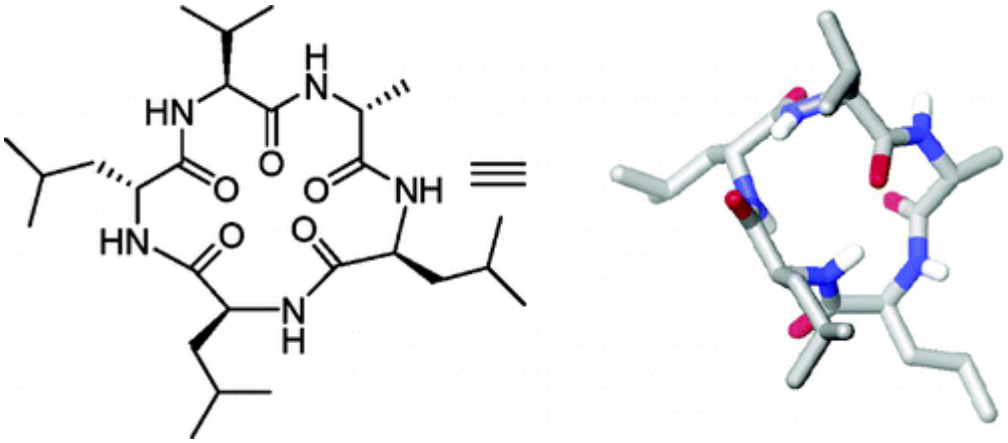


Figure 7.1: Left - structure of **7.1**. Right – Lowest energy conformer of **7.1** from conformational search.

7.2 Modelling

Since attempts to crystallize chrysosporide (compound **7.1**) were unsuccessful, the actual amino acid sequence of **7.1** was established from computer modelling and ROESY NMR studies. A conformational search of cyclopentapeptides, containing L-Val, D-Ala, and all three possible combinations of two L-Leu and one D-Leu, was carried out with MacroModel[5] in combination with the OPLS2003 force field using a GB/SA water solvent model. An ensemble of conformers was collected that lay within 12 kJ of the global minima.

The NOE is a function approximating to r-6 (r = H/H distance), which means for example that a H/H distance of 2 Å will give a much stronger NOE signal than a H/H distance of 3 Å. Boltzmann weighted r-6 total values were calculated for each of the inter-residue H/H distance pairs within each ensemble of conformations within a 12 kJ/mol window. These values were used to compare the ensembles of conformers of the three structures against the experimentally measured NOEs.

In the ROESY spectrum of **7.1** five strong inter-residue NOEs were observed. **Table 7.1** lists the global minima and the Boltzmann-weighted average of H/H distances for each

ensemble, along with the Boltzmann weighted r-6 values for each ensemble. Only cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu) (**7.1**) showed appropriate calculated distance data (relatively large values for the Boltzmann weighted r-6 total values) for the five NH/Hα and NH/NH combinations for which strong NOEs were observed. The other two structures have two or three calculated H/H values, which would not result in strong NOE signals (relatively small values for the Boltzmann weighted r-6 totals in **Table 7.1**). Therefore, the sequence of the isolated product was proposed to be cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu).

Table 7.1. Selected Inter-residue H/H Distancesa in Diastereoisomeric Structures of Chrysosporide (1) Generated by Molecular Modeling

		Hα(Ala)/ NH(Leu1)	NH(Leu1)/ NH(Leu2)	NH(Leu2)/ NH(Leu3)	NH(Val)/ Hα(Leu3)	Hα(Leu2)/ NHLeu3
cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu)	GMDb	2.21	2.58	2.54	2.68	3.23
	BWA c	2.22	2.71	3.04	2.6	2.86
	BWr-6, d	2.5	0.96	0.69	1.31	1.05
cyclo-(L-Val-D-Ala-L-Leu-D-Leu-L-Leu)	GMD	2.52	4.32	4.51	3.36	2.17
	BWA	2.5	4.13	4.46	3.42	2.2
	BWr-6	0.53	0.03	0.02	0.08	1.05
cyclo-(L-Val-D-Ala-D-Leu-L-Leu-L-Leu)	GMD	3.48	4.03	4.47	3.56	2.13
	BWA	3.26	3.42	3.95	3.53	2.65
	BWr-6	0.42	0.35	0.11	0.1	1.13

a Inter-residue NOEs between these protons were experimentally observed.b GMD - Global minima distance (Å).c BWA − Boltzmann-weighted average distance in ensemble (Å).d BWr-6 − Boltzmann-weighted r-6 total (r = H/H distance)

Examination of the low-energy conformers of cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu) as obtained by a conformational search using MacroModel[5] showed intramolecular hydrogen bonding involving NH-Val and NH-Leu2 (**Figure 7.2**). However, this was a supportive, but not definitive in helping to resolve between the LLD, LDL DLL diastereomers, as all of the low-energy conformers of each possible structure formed the two H-bonds with equal frequency.

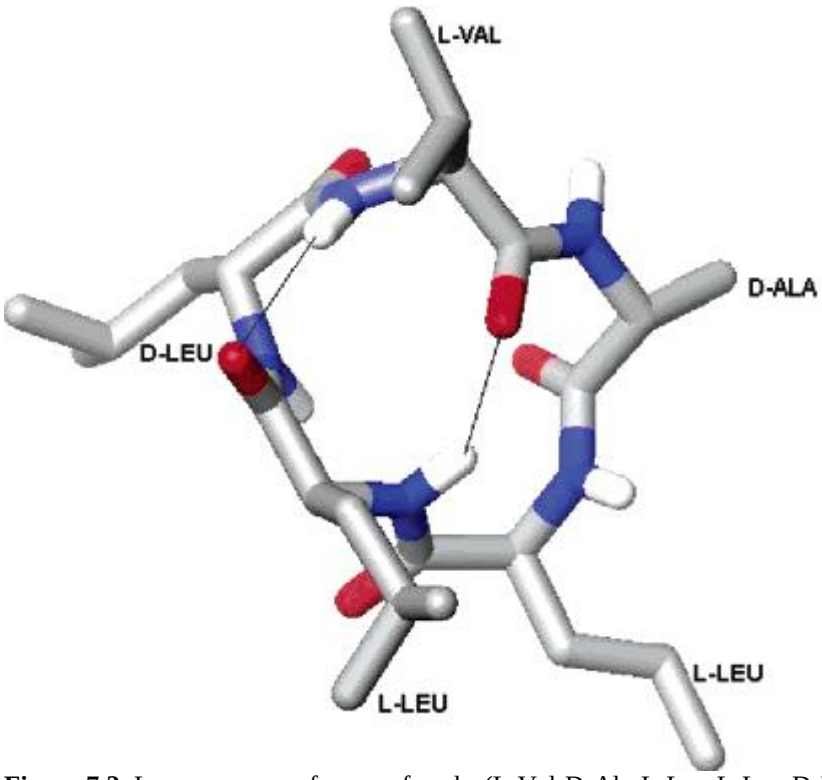


Figure 7.2: Low-energy conformer of cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu) showing intramolecular H-bonds.

The molecular modelling of the three possible LLD, LDL, and DLL diastereomers of the cyclopentapeptides based on the NMR structural analysis were constructed in silico using the Maestro build function of the 2005 Schrödinger molecular modeling suite. A Monte Carlo multiple minimum (MCMM) conformational search of each structure was carried out with MacroModel 9.0 using the OPLS2003 force field with the GB/SA water solvent model. The criteria for convergence of each conformational search were the generation of 3000 starting conformations and a maximum of 5000 iterations in the energy miminization routine (PRCG method) for each conformer, collecting the ensemble of conformers within 12 kJ of the global established using the following expression: $P_\alpha = \frac{\exp[-(E_\alpha/k_B T)]}{\sum_{\alpha}^{N_A} \exp[-(E_\alpha/k_B T)]}$ The five NH/Hα and NH/NH distances, for which strong NOEs were observed, were measured for all conformers within each ensemble. The Boltzmann weighted r-6 total was calculated for these five H/H distances within each ensemble, which gave virtual NOE values. These were used to compare the

modelled data with the experimental data in order to determine the amino acid sequence of chrysosporide (**see Table 7.1**).

7.3 Conclusion

Modelling showed the cyclo-(L-Val-D-Ala-L-Leu-L-Leu-D-Leu) to be the most likely diastereomer as the Boltzmann weighted distance data was calculated to be consistent with the observed NOEs. This was not the case with the other two possible diastereomers.

References

1. Sahr, T.; Ammer, H.; Besl, H.; Fischer, M., *Mycologia* **1999,** *91*, 935-943.
2. Marfey, P., Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Research Communications* **1984,** *49*, 591-596.
3. Mitova, M. I.; Lang, G.; Blunt, J. W.; Cummings, N. J.; Cole, A. L. J.; Robinson, W. T.; Munro, M. H. G., Cladobotric acids A-F: new cytotoxic polyketides from a New Zealand Cladobotryum sp. *Journal of Organic Chemistry* **2006,** *71*, 492-497.
4. Alley, M. C.; Scudiero, D. A.; Monks, A.; Hursey, M. L.; Czerwinski, M.; Fine, D. L.; Abbott, B. J.; Mayo, J. G.; Shoemaker, R. H.; Boyd, M. R., Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. *Cancer research* **1988,** *48*, 589-601.
5. *MacroModel*, version 9.1; Schrödinger: LLC, New York, NY, 2005.

8 Appendix

8.1 Protocol 1: Refinement of the Glide model from 1KXR

The crystal structure of mini μ-calpain (pdb code 1KXR)[1] was prepared by deleting water and ions, and mutation of Ser115->Cys115. This structure was minimised using the OPLS2001 force field with a GB/SA water model over 500 iterations. All residues within a 5 Å distance to the calcium ions of the crystal structure were kept frozen during this minimisation. The RMSD of the minimised structure to the crystal structure was 0.96 Å for the heavy atoms (C, N, O, S). The structure was cleaned up by the Prime[2] preparation and refinement tool. The cysteine sulphur of Cys115 was deprotonated.

8.2 Protocol 2: Refinement of the InducedFit model from 1KXR

The crystal structure of mini μ-calpain (pdb code 1KXR)[1] was prepared by deleting water and ions, and mutation of Ser115->Cys115. This structure was minimised using the OPLS2001 force field with a GB/SA water model over 500 iterations. All residues within a 5 Å distance to the calcium ions of the crystal structure were kept frozen during this minimisation. The RMSD of the minimised structure to the crystal structure was 0.96 Å for the heavy atoms (C, N, O, S). The structure was cleaned up by the Prime[2] preparation and refinement tool.

8.3 Protocol 3: Conformational search methods

Conformational searches were carried out with MacroModel 9.1[3] to generate an ensemble of low energy conformers to establish a suitable starting conformation of each compound for the docking.[3] The searches were conducted with the MCMM method using a GB/SA water model and the OPLS2001 force field, with 3000 steps for the conformational search and up to 5000 iterations for the minimisation of each generated structure. The minimisation was stopped with the default gradient convergence threshold of δ = 0.05 kJ/(mol*Å). The default Polak-Ribiere Conjugate Gradient method was used for all minimisations.

8.4 Protocol 4: Glide docking protocol

Grid generation: The centre of the docking grid was defined as the centroid of the residues Cys115, Gly208, Gly271, and Lys347 and was generated with GLIDE 4.0 using default settings.[4] The midpoint of each docked ligand was set to a 12 x 12 x 12 Å box. Van der Waals radii of the receptor atoms were scaled to 1.0. All other settings were default settings.

Ligand docking: Ligands were docked flexibly in extra precision mode (XP) and with a van der Waals scaling of 0.8. The structure output was set to write out 10 poses per ligand. All other settings were default settings.

8.5 Protocol 5: InducedFit docking protocol

The InducedFit[5] docking script was opened in Maestro.[6]

Glide enclosing box: The centre of the docking grid was defined as the centroid of the residues Cys115, Gly208, Gly271, and Lys347. The size was set at auto.

Step 1: No protein preparation constrained refinement. Remove side chain of Lys347. Extra precision (XP). All other settings at default.

Step 2: Default settings

Step 3: Extra precision (XP). All other settings as default.

8.6 Protocol 6: Computational methods used for diazo- and triazene-dipeptide aldehydes in section 3.4

All molecular modeling experiments were conducted with the Schrödinger suite 2005. Conformational searches on **3.36a-d**, **3.37a/b**, and **3.38a/b** were carried out with MacroModel 9.1 to generate an ensemble of low energy conformers to establish a suitable starting conformation of each compound for the docking.[3] The searches of (*E*)-isomers were conducted with the MCMM method using a GB/SA water model and the OPLS2001 force field, with 3000 steps for the conformational search and up to 5000 iterations for the minimisation of each generated structure. The searches of (*Z*)-isomers were conducted with the MCMM method using a GB/SA water model and the MM2* force field where the C-N=N-C dihedral angle was constrained to 8° with a force constant of 5000 kJ/mol. The minimisation was stopped with the default gradient convergence threshold of $\delta = 0.05$ kJ/(mol*Å). The default Polak-Ribiere Conjugate Gradient method was used for all minimisations. The centre of the docking grid was defined as the centroid of the residues

Cys115, Gly208, and Gly271 and was generated with GLIDE 4.0 using default settings.[4] The centroid of Cys115, Gly208, Gly271, Glu349, and Asn253 was chosen for the docking grid generation and the size of the box was chosen by default. Van der Waals radii of the ligand and the protein atoms were scaled to 0.5. The side chain of Lys347 was removed for the docking and the 20 best poses of this S25 initial docking were retained. All protein residues within a 5 Å of the respective ligand pose were refined with PRIME 1.5, including Lys347.[5] The ligands were redocked with a van der Waals radius of 0.8 to the newly generated protein structures if within 30 kcal/mol of the best protein structure and only if within the top 20 structures. For each of these protein structures, one ligand pose was kept for evaluation.

8.7 Protocol 7: Preparation of dehydroquinase model

The crystal structure of dehydroquinase (pdb code 1HOS) was prepared by deleting all water molecules. The structure was then cleaned up by the Prime preparation and refinement tool. The RMSD of the minimised structure to the crystal structure was 0.03 Å for the heavy atoms (C, N, O, S).

8.8 Protocol 8: Glide docking protocol for dehydroquinase

The receptor was scaled to 1.0 and the ligand identified. The centre of the docking grid was deemed to be the centre of the workspace ligand and the size of the ligands to be docked was those similar in size to the workspace ligand. Five H-bond constraints were identified; the H-bonds of which there are two to Ser103, one to Ile102, and two to Asn75.

8.9 Protocol 9: Bodipy assay

The biological activity of the calpain inhibitors synthesized in this thesis was determined by measuring the inhibition constants (IC50) in an *in vitro* assay. IC_{50} indicates the concentration of inhibitor required to decrease the activity of the enzyme by 50%.

Collaboration with Lincoln University allowed access to the materials required for the biological assay of cysteine protease inhibitors. The testing was performed by Dr Janna Nikkel and others in our research group. An established assay protocol for the determination of m-calpain inhibition based on a fluorogenic methodology used by Thompson et al[7] was used. This assay protocol utilises casein, a water soluble protein, labelled with the fluorophore 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-*s*indacene-3-propionic acid (BODIPY). Fluorescence increases as proteolysis of the substrate (casein) occurs. Without proteolysis, fluorescence is not observed as adjacent intramolecular fluorophores cause auto-quenching of fluorescence. Thus the inhibitory activity of a proposed inhibitor can be measured by calculating the change in fluorescence over a known period of time. m-Calpain and μ-calpain were partially purified from sheep lung by ion-exchange chromatography and diluted to give a linear response over the course of the assay.

References

1. Moldoveanu, T.; Hosfield, C. M.; Lim, D.; Elce, J. S.; Jia, Z.; Davies, P. L., A Ca2+ Switch Aligns the Active Site of Calpain. *Cell* **2002,** *108,* 649-660.
2. *Prime*, version 1.5; Schrödinger: LLC, New York, NY, 2005.
3. *MacroModel*, version 9.1; Schrödinger: LLC, New York, NY, 2005.
4. *Glide*, version 4.0; Schrödinger: LLC, New York, NY, 2005.

5. *Schrödinger Suite 2006 Induced Fit Docking Protocol*, Glide version 4.0, Prime version 1.5; Schrödinger: LLC, New York, NY, 2005.

6. *Maestro*, version 7.5; Schrödinger: LLC, New York, NY, 2005.

7. Thompson, V. F.; Saldana, S.; Cong, J.; Goll, D. E., A BODIPY fluorescent microplate assay for measuring activity of calpains and other proteases. *Anal. Biochem.* **2000,** *279*, 170–178.